PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G01N 33/50, 33/68

A1

(11) International Publication Number: WO 99/28744

(43) International Publication Date: 10 June 1999 (10.06.99)

(21) International Application Number: PCT/EP98/07656

(22) International Filing Date: 27 November 1998 (27.11.98)

(30) Priority Data:

97120879.8	27 November 1997 (27.11.97)	EP
97120867.3	27 November 1997 (27.11.97)	EP
97120880.6	27 November 1997 (27.11.97)	EP

(71) Applicant *(for all designated States except US)*: MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN E.V. [DE/DE]; Berlin (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WANKER, Erich [AT/DE]; Leichardstrasse 61, D-14195 Berlin (DE). BANCROFT, David [GB/DE]; Pasinger Strasse 11, D-82152 Martinsried (DE). LEHRACH, Hans [AT/DE]; Lützelsteiner Weg 50, D-14195 Berlin (DE). WEDEMEYER, Niels [DE/DE]; Clayallee 228, D-14195 Berlin (DE).

(74) Agents: KOLB, Helga et al.; Hoffmann . Eitle, Arabellastrasse 4, D-81925 München (DE).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: IDENTIFICATION AND CHARACTERIZATION OF INTERACTING MOLECULES

(57) Abstract

The present invention relates to an improved method for the identification and optionally the characterization of interacting molecules designed to detect positive clones from the rather large numbers of false positive clones isolated by two-hybrid systems. The method of the invention relies on a novel combination of selection steps used to detect clones that express interacting molecules from false positive clones. The present invention further relates to a kit useful for carrying out the method of the invention. The present invention provides for parallel, high-throughput or automated interaction screens for the reliable identification of interacting molecules.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

IDENTIFICATION AND CHARACTERIZATION OF INTERACTING MOLECULES

The present invention relates to an improved method for the identification and optionally the characterization of interacting molecules designed to detect positive clones from the rather large numbers of false positive clones isolated by two-hybrid systems. The method of the invention relies on a novel combination of selection steps used to detect clones that express interacting molecules from false positive clones. The present invention further relates to a kit useful for carrying out the method of the invention. The present invention provides for parallel, high-throughput or automated interaction screens for the reliable identification of interacting molecules.

Protein-protein interactions are essential for nearly all biological processes like replication, transcription, secretion, signal transduction and metabolism. Classical methods for identifying such interactions like co-immunoprecipitation or cross-linking are not available for all proteins or may not be sufficiently sensitive. Said methods further have the disadvantage that only by a great deal of energy, potentially interacting partners and corresponding nucleic acid fragments or sequences may be identified. Usually, this is effected by protein sequencing or production of antibodies, followed by the screening of an expression-library.

An important development for the convenient identification of protein-protein interactions was the yeast two-hybrid (2H) system presented by Fields and Song (1989).This genetic procedure not only allows the rapid demonstration of in vivo interactions, but also the simple isolation of corresponding nucleic acid sequences encoding for the interacting partners.

The yeast two-hybrid system makes use of the features of a wide variety of eukaryotic transcription factors which carry two separable functional domains: one DNA binding domain as well as a second domain which activates the RNA-polymerase complex (activation domain). In the classical 2H system a so-called "bait" protein comprising of a DNA binding domain (GAL4bd or lex A) and a protein of interest "X" are expressed as a fusion protein in yeast. The same yeast cell also simultaneously expresses a so called "fish" protein comprising of an activation domain (GAL4ad or VP16) and a protein "Y". Upon the interaction of a bait protein with a fish protein, the DNA binding and activation domains of the fusion proteins are brought into close proximity and the resulting protein complex triggers the expression of the reporter genes, for example, HIS3 or lacZ. Said expression can be easily monitored by cultivation of the yeast cells on selective medium without histidine as well as upon the activation of the lacZ gene. The genetic sequence encoding, for example, an unknown fish protein, may easily be identified by isolating the corresponding plasmid and subsequent sequence analysis. Meanwhile, a number of variants of the 2H system have been developed. The most important of those are the "one hybrid" system for the identification of promoter binding proteins and the "tri-hybrid" system for the identification of RNA-protein-interactions (Li and Herskowitz, 1993; SenGupta et al., 1996; Putz et al., 1996). It is understood in the art that to identify, detect or assay the variety of interactions found in biological systems, different 2H systems must be employed. Indeed, other 2H technologies have been developed to enable protein-protein interactions to be investigated in other organisms and/or different cell compartments. For example, in mammalin cells (Rossi et al, 1997; PNAS 94:8405-8410), in bacterial cells (Karimova et al., 1998; PNAS 95:5752-5756), in the cytoplasm of yeast cells (Johnsson & Varshavsky; 1996 US5503977) and in the periplasmic space of yeast cells (Fowlkes et al., 1998; US 5789184).

These 2H systems for the identification of protein-protein-interaction, have, until today, only been carried out on a laboratory scale. The various steps of these systems need to be conducted serially. They are, therefore, quite time consuming. As a consequence, these 2H systems have so far proven unsuitable for the analysis of eukaryotic library vs library screens to investigate protein-protein networks. Although recent developments have taken into account these disadvantages (Bartel et al.,1996), a successful large scale search of interacting proteins, for example on the basis of a eukaryotic library vs. library screen, has not been reported. More importantly also, 2H systems suffer from the serious drawback that many false-positive clones not representing any interactions between binding partners are isolated. This is particularly inconvenient in cases where large numbers of clones are to be analyzed because in the case of a eukaryotic library vs library screen it is typical that several hundreds of thousands of clones have to be analyzed for the investigation of protein-protein networks.

The technical problem underlying the present invention was therefore to overcome these prior art difficulties and to furnish a system that reliably produces clones that express interacting molecules. This system should, moreover, be suitable for large-scale library vs library screens using a parallel, high-throughput or automated approach.

The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to a method for the identification of at least one member of a pair or complex of interacting molecules, comprising:

(a) providing host cells containing at least two genetic elements with different selectable and counterselectable markers, said genetic elements each comprising genetic information specifying one of said members, said host

cells further carrying a readout system that is activated upon the interaction of said molecules;

(b) allowing at least one interaction, if any, to occur;

(c) selecting for said interaction by transferring progeny of said host cells to

(ca) at least two different selective media, wherein each of said selective media allows growth of said host cells only in the absence of at least one of said counterselectable markers and in the presence of a selectable marker; and

(cb) a further selective medium that allows identification of said host cells only on the activation of said readout system ;

(d) identifying host cells containing interacting molecules that

(da) do not activate said readout system on any of said selective media specified in (ca); and

(db) activate the readout system on said selective medium specified in (cb); and

(e) identifying at least one member of said pair or complex of interacting molecules.

Preferably, said interaction is a specific interaction.

The terms "identification" and "identifying", as used in accordance with the present invention, relate to the ability of the person skilled in the art to detect positive clones that express interacting molecules from false positive clones due to the activation of the readout system on the selective media and optionally additionally to characterize at least one of said interacting molecules by one or a set of unambiguous features. Preferably, said molecules are characterized by the DNA sequence encoding them, upon nucleic acid hybridization or isolation and sequencing of the respective DNA molecules. Alternatively and less preferred, said molecules may be characterized by different features such as molecular weight, isoelectric point and, in the case of proteins, the N-terminal

amino acid sequence etc. Methods for determining such parameters are well known in the art.

Preferably, said members specified by said genetic elements are connected to a further entity that will upon the interaction activate or contribute to the activation of said read out system. It is further preferred that said entity is conserved for each type of genetic element and that different types of genetic elements comprise different entities. It is additionally preferred that said member of said pair or complex of interacting molecules forms, when transcribed as RNA from said genetic element, an RNA transcript fused with RNA specifying said entity. Most preferably, said fused RNA transcript is translated to form a fusion protein comprising said member fused to said entity. As will be elaborated further herein below, said entity may be in one type of genetic element a DNA sequence encoding a DNA-binding domain and in a different type of genetic element a transactivating protein domain. Preferably, said genetic elements are vectors such as plasmids. Alternatively, interaction between two fusion proteins may result in a functional entity with reconstituted enzymatic activity, for example the bacterial chloramphenicol acetyltransferase protein (CAT) (Seed & Sheen, 1988 Gene 67:271-277). The at least two genetic elements comprised in said host cell are preferentially vectors from a library such as a cDNA or genomic library. Thus, the method of the invention allows the screening of a variety of host cells wherein the vector portion of said genetic elements is preferably the same for each type of genetic element whereas the potentially interacting molecules are representatives of a library and, thus, as a rule and in case that the library has not been amplified, may differ in each host cell. In this connection the term "type of genetic element" refers to an element characterized by comprising the same entity, selectable and counterselectable markers.

Preferably, the "interaction" of said molecules is specific and characterized by a high binding constant. However, the term "interaction" may also refer to a binding between molecules with a lower binding constant which, however, must be sufficient to activate the readout system. The interaction that is detectable by the method of the invention preferably leads to the formation of a functional entity having a biological, physical or chemical activity which was not present in said host cell before said interaction occurred.

Said interaction may lead to the formation of a functional transcriptional activator comprising a DNA-binding and a transactivating protein domain and which is capable of activating a responsive moiety that drives the activation of said readout system. For example, said moiety may be a promoter.

Alternatively, said interaction may lead to a detectable fluorescence resonance energy transfer obtained by the interaction of fusion proteins containing, for example, the GFP type a and GFP type b fluorescent proteins (Cubbitt et al., 1995; Heim & Tsien, Curr Biol. 1996 6:178-182). Said interaction may also alternatively lead to the reconstitution of a functional enzyme, for example ß-galacotsidase (Rossi et al., 1997) or adenylate cyclase (Karimova et al., 1998). These embodiments will be preferred for the study of interactions in host-cell types other than yeast.
In a further embodiment, said interaction may lead to a detectable modification of a substrate by an enzyme such as a color reaction obtained by the cleavage of a propeptide by an enzyme. In all these embodiments of the invention, it is understood that the interacting molecules are preferably directly fused to the molecules driving the readout system.

The term "growth" on selective media "in the absence of at least one of said counter-selectable markers" refers to the fact that a population of host cells containing at least one

of genetic elements is placed on said selective media but only those progeny of the host cells in the overall population that have lost the relevant genetic element are able to grow. For example, when a yeast strain which is resistant to the drug canavanine (can^r) and which also contains a plasmid carrying the wild-type CAN1 gene (Hoffmann, 1985) is placed on a selective medium containing canavanine, only those progeny of the yeast strain that have lost the plasmid carrying the CAN1 gene are able to grow, because this gene confers sensitivity to canavanine in yeast cells.

With reference to step (ca), it should be noted that each of the at least two selective media would comprise at least one counterselectable compound such as cycloheximide wherein the counterselectable compound would be different in the different selective media; they would further typically lack a compound complementing for an auxotrophic marker or comprise an antibiotic. The compound or antibiotic may be the same for the various selective media. Preferably, at least one is different.

The method of the present invention provides a highly effective tool for selecting against false positive clones that have proven to dramatically reduce the overall usefulness of the two-hybrid system. For example, by inclusion of a marker counterselecting for the absence of a genetic element that specifies one of a pair of the potentially interacting partners, clones that will grow and therefore only carry the second genetic element specifying the second partner can now be tested for the activation of the readout system. If the clone containing only the fusion protein encoded by the second genetic element activates the readout system in the absence of the other genetic element, then it will be classified as a false positive. By counterselecting for the absence of the second genetic element, the same test is applied to the first genetic element. Thus, only clones that activate the readout system in the presence of both or all genetic elements, but do

not activate the read out system when either of the genetic elements is lost are classified as positives.

The advantages associated with the method of the invention have a significant impact in particular on the number of clones that express potentially interacting partners that can conveniently be analyzed. For example, even work on the laboratory scale will be more effective since positive clones that express interacting partners can be easily and unambiguously discriminated from false positive clones without the generation of additional strains. In contrast, to detect false positive clones using the state of the art yeast two-hybrid system, plasmids that encode fish proteins usually need to be isolated and retransformed into yeast cells harboring plasmids that encode unrelated bait proteins. Further, the enormous number of false positive clones that would be isolated when using the classical two-hybrid system on a large scale, yet are discriminated by the method of this invention no longer precludes an effective high through-put analysis of clones. In the long run, it is expected that the method of the present invention is especially advantageous for a high throughput analysis of a large number of yeast clones containing interacting molecules since many specific interactions and the individual members of these interactions can be identified in a parallel and automated approach.

Some investigators have noted the problem of identifying false positive clones when applying the yeast two-hybrid system in the past. Bartel et al. (1996) described a method for the elimination of false positives by replica plating clones that express one fusion protein from SD-leu and SD-trp plates, to SD-his plates. Clones that showed growth on the SD-his plates where identified as false positives and were subsequently not used for interaction mating. The disadvantage of this method is that the procedure is labor intensive because yeast strains expressing the fish proteins, the bait proteins and the potentially interacting fish and bait proteins all must be

generated and analyzed. The use of the counterselectable system described in this invention has the advantage that only one strain which expresses the potentially interacting fusion proteins is generated and must be analyzed.

Other strategies have been proposed to eliminate false positive clones from 2H systems (Vidal et al., 1996a; Nandabalan et al., 1997). However, these systems all require that the readout system that is assayed for activity comprises at least one reporter gene that is transcribed on reconstitution of DNA binding and transactivating fusion proteins. Indeed, although mostly claiming to be applicable to all types of cells, these systems have been designed towards the specific biological properties of the yeast two-hybrid system. The method if invention described herein is not limited to eliminating false positive clones expressing single DNA binding or activation domain fusion proteins that can activate the reporter system. On the contrary, it can be uses to eliminate false positive clones in 2H systems other than yeast two-hybrid, which is of advantage when interaction screens are conducted in for example, other host-cell types.

A schematic overview of one embodiment of the method of the invention is provided in Figure 6. For the parallel analysis of a network of protein-protein interactions with the method of the invention, a library of plasmid constructs that express DNA binding domain and activation domain fusion proteins is provided. These libraries may consist of specific DNA fragments or a multitude of unknown DNA fragments ligated into the improved binding domain and activating domain plasmids of the invention containing different selectable and counterselectable markers. Both libraries are combined within yeast cells by transformation or interaction mating, and yeast strains that express potentially interacting proteins are selected on selective medium lacking histidine. The selective markers TRP1 and LEU2 maintain the plasmids in yeast strains grown on selective media, whereas CAN1 and CYH2 specify the

counter-selectable markers that select for the loss of each plasmid. HIS3 and lacZ represent selectable markers integrated into the yeast genome, which are expressed on activation by interacting fusion proteins.

The readout system is, in the present case, both growth on medium lacking histidine and enzymatic activity of ß-galactosidase which can be subsequently screened. It is to be understood, however, that the readout system may rely on only one marker such as HIS3. Yet, the combination of two components that constitute the readout system in many cases allows a more ready interpretation of results, in particular if one of the components, when activated, effects a change in color. A colony picking robot is used to pick the resulting yeast colonies into individual wells of 384-well microtiter plates containing selective medium lacking histidine, and the resulting plates are incubated at 30°C to allow cell growth. The interaction library contained in microtiter plates can be optionally replicated and stored. The resulting interaction library is investigated to detect positive clones that express interacting proteins and discriminate them from false positive clones using the method of the invention. Using a spotting robot, cells are transferred to replica membranes which are subsequently placed onto one each of the selective media SD-leu-trp-his, SD-leu+CAN and SD-trp+CHX. After incubation on the selective plates, the clones grown on the membranes are subjected to a ß-Gal assay and a digital image from each membrane is obtained with a CCD camera which is then stored on computer. Using digital image processing and analysis clones that express interacting fusion proteins can be identified by considering the pattern of ß-Gal activity from clones grown on the various selective media. The individual members comprising interactions can then be identified by one or more techniques, including PCR, sequencing, hybridization, oligofingerprinting or antibody reactions. An actual experiment carried out along the schematic route presented in Figure 6 is shown in Figure 5 to Figure 22.

The genetic elements specified here and above may further and advantageously be equipped with at least two different selection markers functional in bacteria such as E.coli. Such selection markers, for example aphA (Pansegrau et al., 1987) or bla allow the easy separation of said genetic elements upon retransformation into E.coli strains.

In a preferred embodiment of the method of the present invention said pair or complex of interacting molecules is selected from the group consisting of RNA-RNA, RNA-DNA, RNA-protein, DNA-DNA, DNA-protein, protein-protein, protein-peptide, or peptide-peptide interactions.

Accordingly, the method of the invention is applicable in a wide range of biological interactions. For example, the invention will be useful in identifying peptide-protein or peptide-peptide-interactions by employing synthetic peptide libraries (Yang et al., 1995).

Two applications of interests are the application of a large scale two-hybrid system for the detection of protein-protein interactions involved in medically relevant pathways which may be useful as diagnostic or therapeutic targets for the treatment of disease, and a large scale tri-hybrid system which is one example of said complex of interacting molecules mentioned herein above for the identification of, for example, novel post-transciptional regulators and their binding sites (SenGupta et al., 1996; Putz et al., 1996). In this regard it should be noted that a complex, in accordance with the invention may comprise more than three interacting molecules. Furthermore, such a complex may be composed of biologically or chemically different members. For example, to identify interacting RNA binding proteins and RNA molecules, a plasmid expressing a LexA-HIV-1Rev protein, a plasmid transcribing an RNA sequence in fusion with the Rev responsive element and a plasmid expressing a potentially RNA-interacting protein in

fusion with an activation domain may be present in one cell. The plasmids encoding the RNA fusion molecule and the activation domain fusion protein must contain different selectable and counterselectable markers according to the method of the invention. If the RNA fusion molecule interacts with the respective two fusion proteins, the readout system is activated. To test whether the RNA fusion molecule or the activation domain fusion protein interact, the method of the invention is used to investigate the activation of the readout system in the absence of either of these fusion molecules.

In a further preferred embodiment, said genetic elements are plasmids, artificial chromosomes, viruses or other extrachromosomal elements.

Whereas it is preferred, due to the easy handling, to employ plasmids that specify the genetic elements in accordance with the present invention, the persons skilled in the art will be able to devise other systems that carry said genetic elements and that are identified above.

In an additional preferred embodiment, said readout system is a detectable protein. A number of readout systems are known in the art and may, if necessary, be adapted to be useful in the method of the invention.

Most preferably, said detectable protein is that encoded by the gene lacZ, HIS3, URA3, LYS2, sacB or HPRT, respectively. As is well known in the art, the expression of the ß-gal enzyme in yeast can be used for the formation of a detectable blue colony after incubation in X-Gal solution. Of course, the method of the invention is not restricted for use of only one readout system. On the contrary, if desired, a number of such readout systems may be combined. Said combination of a number of readout systems is, in accordance with the present invention, also comprised by the term "readout system". Such a

combination will provide an additional safe guard for the identification of clones containing interacting partners.

Although the two-hybrid system has been developed in yeast, the method of the invention can be carried out in a variety of host systems. Preferred of those are yeast cells, bacterial cells (Karimova et al., 1998), mammalian cells (Wu et al. 1996, Rossi et al., 1997), insect cells or plant cells. Preferably, the bacterial cells are E. coli cells.
Of course, the genetic elements may be engineered and prepared in one host organism and then, e.g. by employing shuttle vectors, be transferred to a different host organism where it is employed in the method of the invention.

In another preferred embodiment, the method of the present invention comprises transforming or transfecting said host cell with at least one of said genetic elements prior to step (a).

Whereas the person skilled in the art may initiate the identification method of the invention starting from fully transformed or transfected host cells, he may wish to first generate such host cells in accordance with the aim of his research or commercial interest. For example, he may wish to generate a certain type of library first that he intends to screen against a second library already present in said host cells. Alternatively, he may have in mind to generate two or more different libraries that he wants to screen against each other. In this case, he would need to first transform said host cells, simultaneously or successively, with both or all types of genetic elements.

In another preferred embodiment, said host cells with said genetic elements are generated by cell fusion, conjugation or interaction mating.

The biological principal of counter-selection referred to above is well known in the art. Accordingly, the person skilled in the art may chose from a variety of such counter-selectable markers. Preferably, said markers are CAN1, CYH2, LYS2, URA3, HRPT or sacB.

It is further preferred in accordance with the present invention that said selectable markers are auxotrophic or antibiotic markers.

It is important to note that some of the markers that are used as a readout system, may also be used as selectable markers. It is further important to note that one and the same marker can not be used as selectable marker and as part of the readout system at the same time.

Most preferably, said auxotrophic or antibiotic markers are selected from LEU2, TRP1, URA3, HIS3, ADE2, LYS2 and Zeocin.

Planning of experiments may require that the test for interaction need not be done immediately after the provision of host cells and, possibly, the occurrence of the interactions. In such cases, the researcher may wish to store the transformed host cells for further use. Accordingly, a further preferred embodiment of the invention relates to a method wherein progeny of host cells obtained in step (b) are transferred to a storage compartment.

In particular in cases where a large number of clones is to be analyzed, said transfer is advantageously effected or assisted by automation or a picking robot. Naturally, other automation or robot systems that reliably pick progeny of said host cells into predetermined arrays in the storage compartments may also be employed.

The host cells will, in this embodiment, be propagated in said storage compartment and provide further progeny for the

additional tests. Preferably, replicas of said storage compartment maintaining the array of clones are set up. Said storage compartments comprising the transformed host cells and the appropriate media may be maintained in accordance with conventional cultivation protocols. Alternatively, said storage compartments may comprise an anti-freeze agent and therefore be appropriate for storage in a deep-freezer. This embodiment is particularly useful when the evaluation of potential interacting partners is to be postponed. As is well known in the art, frozen host cells may easily be recovered upon thawing and further tested in accordance with the invention. Most preferably, said anti-freeze agent is glycerol which is preferably present in said media in an amount of 3 - 25% (vol/vol).

In a further particularly preferred embodiment of the method of the invention, said storage compartment is a microtiter plate. Most preferably, said microtiter plate comprises 384 wells. Microtiter plates have the particular advantage of providing a pre-fixed array that allows the easy replicating of clones and furthermore the unambiguous identification and assignment of clones throughout the various steps of the experiment. The 384 well microtiter plate is, due to its comparatively small size and large number of compartments, particularly suitable for experiments where large numbers of clones need to be screened.

Depending on the design of the experiment, the host cells may be grown in the storage compartment such as the above microtiter plate to logarithmic or stationary phase. Growth conditions may be established by the person skilled in the art according to conventional procedures. Cell growth is usually performed between 15 and 45 degrees Celsius.

Transfer of said host cells in step (c) is made or assisted by automation, by using a spotting robot or by using a pipetting or micropipetting device. How such a spotting robot may be

devised and equipped is, for example, described in Lehrach et al. (1997). Naturally, other automation or robotic systems that reliably create ordered arrays of clones may also be employed.

Most preferably, said transfer is made to a planar carrier which is subsequently placed on the at least three selective media as specified in steps (ca) and (cb). Alternatively, said transfer of said host cells may be made to the planar carrier already placed on the selective media or said transfer may be made directly to the selective media.

Most advantageously, said transfer is effected in a regular grid pattern at densities of 1 to 1000 clones per square centimeter. The progeny of said host cells may be transferred to a variety of planar carriers. Most preferred is a membrane which may, for example, be manufactured from nylon, nitro-cellulose or PVDF.

The selective media used for growth of appropriate clones may be in liquid or in solid form. Preferably, said selective media when used in conjunction with a spotting robot and membranes as planar carriers are solidified with agar on which said spotted membranes are subsequently placed. Alternatively, and also preferably, said selective media when in liquid form are held within microtiter plates and said transfer is made by replication.

Referring now to the step (d) of the method of the invention, the readout system can be analyzed by a variety of means. For example, it can be analyzed by visual inspection, radioactive, chemiluminescent, fluorescent, photometric, spectrometric, infra red, colourimetric or resonant detection.

Preferably, said identification of host cells that express interacting fusion proteins is effected by visual means from consideration of the activation state of said readout system

of clones grown on the at least three selective media as specified in steps (ca) and (cb).

Also preferably, said identification of host cells that express interacting fusion proteins in step (d) is effected or assisted by digital image storage, analysis or processing. In this embodiment, positive clones which are preferably arrayed on a planar carrier such as a membrane are identified by comparison of digital images obtained from the membrane after activation of said readout system on said selective media specified in (ca) and (cb).

Most preferably, the identity of positive host cells and false positive host cells are stored on computer, for example within a relational database.

Identification of the at least one member of the pair or complex of interacting molecules may be effected by a variety of means. For example, molecules can be characterized by nucleic acid hybridization, oligonucleotide hybridization, nucleic acid or protein sequencing, restriction digestion, spectrometry or antibody reaction. Once the first member of an interaction has been identified, the second member or further members can also be identified by any of the above methods. Preferably the identification of at least one member of an interaction is effected by nucleic acid hybridization, antibody binding or nucleic acid sequencing.

If nucleic acid hybridization is to be carried out, the nucleic acid molecules comprised in the host cell and encoding for at least one of the interacting molecules is preferably affixed to a planar carrier. As is well known in the art, said planar carrier to which said nucleic acid may be affixed, can be for example, a Nylon-, nitrocellusose- or PVDF membrane, glass or silica substrates (DeRisi et al. 1996; Lockhart et al. 1996). Said host cells containing said nucleic acid may be transferred to said planar carrier and subsequently lysed on

the carrier and the nucleic acid released by said lysis is affixed to the same position by appropriate treatment. Alternatively, progeny of the host cells may be lysed in a storage compartment and the crude or purified nucleic acid obtained is then transferred and subsequently affixed to said planar carrier. Advantageously, said nucleic acids are amplified by PCR prior to transfer to the planar carrier. Most preferably said nucleic acid is affixed in a regular grid pattern in parallel with additional nucleic acids representing different genetic elements encoding interacting molecules. As is well known in the art, such regular grid patterns may be at densities of between 1 and 50 000 elements per square centimeter and can be made by a variety of methods. Preferably, said regular patterns are constructed using automation or a spotting robot such as described in Lehrach et al. (1997) and Maier et al. (1997) and furnished with defined spotting patterns, barcode reading and data recording abilities. Thus it is possible to correctly and unambiguously return to stored host cells containing said nucleic acid from a given spotted position on the planar carrier. Also preferably, said regular grid patterns may be made by pipetting systems, or by microarraying technologies as described by Shalon et al. (1996), Schober et al (1993) or Lockart et al. (1996). Identification is, again, advantageously effected by nucleic acid hybridization.

Using a detectable nucleic acid probe of interest, homologous nucleic acids which are affixed on the planar carrier can be identified by hybridization. From the spotted position of said homologous identified nucleic acid on the planar carrier, the corresponding host cell in the storage compartment can be identified which contains both or all members of the interaction. The for example second member of the interaction can now be identified by any of the above methods. For example, by use of a radioactively labeled Ras probe, homologous nucleic acids on the planar carrier can be identified by hybridization. The Ras interacting proteins can

now be identified from the corresponding host cell that contains both the first genetic element homologous to the Ras probe and the second genetic element encoding for these Ras interacting proteins.

If multiple oligonucleotide hybridizations are carried out on the nucleic acids affixed to the planar carrier, oligofingerprints of all genetic elements encoding the interacting proteins can be obtained. These oligofingerprints can be used to identify all members of the interactions or those members that belong to specific gene families, as described in Maier et al. (1997).

Advantageously, the nucleic acid molecules that encode the interacting proteins are, prior to identification such as by DNA sequencing, amplified by PCR or in said genetic elements in host cells and preferable in *E. coli*. Amplification of said genetic elements is conducted by multiplication of the *E. coli* cells and isolation of said genetic elements. Methods of identifying the nucleic acids that encode interacting proteins by DNA sequencing and analysis are well known in the art. By amplifying and sequencing the nucleic acids that encode for both or all members of an interaction from the same clone, the identity of both or all members of the interaction can be determined.

If a specific antibody is to be used to determine whether a protein of interest is expressed as a fusion protein within an interaction library, it is advantageous to affix all fusion proteins expressed from the interaction library on to a planar carrier. For example, clones of the interaction library that express fusion proteins can be transferred to a planar carrier using a spotting robot as described in Lehrach et al (1997). The clones are subsequently lysed on the carrier and released proteins are affixed onto the same position. Using, for example, an anti-HIP1-antibody (Wanker et al. 1997), clones from the interaction library that contain HIP1 fusion proteins

and an unknown interacting fusion protein can be identified. The unknown member of the interacting pair of molecules can now be identified from the corresponding host cell by any of the above methods. The antibodies used as probes may be directly detectably labeled. Alternatively, said antibodies may be detected by a secondary probe or antibody which may be specific for the primary antibody. Various alternative embodiments using, for example, tertiary antibodies may be devised by the person skilled in the art on the basis of his common knowledge.

Most advantageously, when said identification of members comprising an interaction is effected using said regular grids, a digital image of the planar carrier after hybridization or antibody reaction is obtained and analysis is effected by digital image storage, processing or analysis using an automated or semi-automated image analysis system, such as described in Lehrach et al. (1997).

Most preferably, the information comprising the identity of the host cell and the identity of the interacting molecules expressed by the genetic elements contained within the host cell are stored on a computer, for example within a relational database.

In accordance with the present invention, it is additionally preferred prior to step (a) that a preselection against clones that express a single molecule able to activate the readout system is carried out on culture media comprising a counterselective compound, for example 5-fluoro orotic acid, canavanine, cycloheximide or α-amino-adipate .

In this embodiment, for example, the URA3 gene is incorporated as a component of the readout system. Clones containing only one of said genetic elements are placed on a selective medium comprising 5-fluoro orotic acid (5-FOA). In the case that

clones that express a single molecule able to activate the readout system, 5-FOA is converted into the toxic 5-fluorouracil. Accordingly, host cells containing auto-activating molecules will die on the selective medium containing 5-FOA.

It is further important to note that the marker used for said preselection cannot be used as a selectable or counterselectable marker at the same time.

The present invention also relates to a method for the production of a pharmaceutical composition comprising formulation said at least one member of the interacting molecules identified by the method of the invention in a pharmaceutically acceptable form.

Said pharmaceutical composition comprises at least one of the aforementioned compounds identified by the method of the invention, either alone or in combination, and optionally a pharmaceutically acceptable carrier or exipient. Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Compositions comprising such carriers can be formulated by conventional methods. These pharmaceutical compositions can be administered to subject in need thereof at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g., by intravenous, intraperitoneal, subcutaneous, intramuscular, topical or intradermal administration. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Dosages will vary but a preferred dosage for intravenous

administration of DNA is from approximately 10^6 to 10^{22} copies of the nucleic acid molecule. Proteins or peptides may be administered in the range of 0,1ng to 10mg per kg of body weight. The compositions of the invention may be administered locally or systematically. Administration will generally be parenterally, e.g., intravenously; DNA may also be administered directly to the target site, e.g., by biolistic delivery to an internal or external target site or by catheter to a site in an artery.

The present invention further relates to a method for the production of a pharmaceutical composition comprising formulating an inhibitor of the interaction of the interacting molecules identified by the method of the invention in a pharmaceutically acceptable form.

The inhibitor may be identified according to conventional protocols. Additionally, molecules that inhibit existing protein-protein interactions can be isolated with the yeast two-hybrid system using the URA3 readout system. Yeast cells that express interacting GAL4ad and LexA fusion proteins which activate the URA3 readout system are unable to grow on selective medium containing 5-FOA. However, when an additional molecule is present in these cells which disrupts the interaction of the fusion proteins the URA3 readout system is not activated and the yeast cells can grow on selective medium containing 5-FOA. Using this method potential inhibitors of a protein-protein interaction can be isolated from a library comprising these inhibitors. Systems corresponding to the URA3 system may be devised by the person skilled in the art on the basis of the teachings of the present invention and are also comprised thereby.

Also, the present invention relates to a method for the production of a pharmaceutical composition comprising identifying a further molecule in a cascade of interacting molecules, of which the at least one member of interacting

molecules identified by any of the above methods is a part of or identifying an inhibitor of said further molecule.

Once at least one member of the interacting molecules has been identified, it is reasonable to expect that said member is a part of a biological cascade. Identification of additional members of said cascade can be effected either by applying the method of the present invention or by applying conventional methods. Also, inhibitors of said further members can be identified and can be formulated into pharmaceutical compositions.

The present invention relates further to a kit comprising at least one of the following:

(f) host cells as identified in any of the preceding claims and at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules containing a counterselectable marker and specified herein above;

(g) host cells as identified in any of the preceding claims and at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified herein above;

(h) at least one genetic element comprising said genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified herein above;

(i) at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified herein above;

(j) host cells comprising at least one and preferably at least two of said genetic elements specified in (h) or (i);

(k) at least one planar carrier carrying nucleic acid or protein from said host cells comprising at least one member of said genetic elements specified herein above wherein said nucleic acid or protein is affixed to said carrier in grid form and optionally solutions to effect hybridization or binding of nucleic acid probes or proteins to said molecules affixed to said grid;

(l) at least one storage compartment, planar carrier or computer disc comprising or/and characterizing genetic elements, host cells, storage compartments or carriers identified in any of (f) to (k); and/or

(m) at least one yeast strain comprising a *can*1 and a *cyh*2 mutation.

Preferably, said kit comprises or also comprises at least one storage compartment containing the host cells of (f), (g) or (j) and/or comprises or also comprises at least one storage compartment containing said genetic information or said potentially interacting molecules encoded by said genetic information as specified in (f) or (h).

The present invention also relates to the use of any of the yeast strains described herein above and in the appended examples for the identification of at least one member of a pair of potentially interacting molecules.

Advantageously, those molecules identified by the method of the present invention as interacting with many different molecules can be recorded. This information can reduce the work needed to further characterise particular interactions since those interactions comprising of a molecule found to

interact with many other molecules within a 2H system may be suspected of being artifactual (Bartel et al., 1993).

Preferably, the data obtained by using the method of the present invention can be accessed through the use of software tools or graphical interfaces that enable to easily query the established interaction network with a biological question or to develop the established network by the addition of further data.
Accordingly, the present invention further relates to a computer implemented method for storing and analysing data relating to potential members of at least one pair or complex of interacting molecules encoded by nucleic acids originating from biological samples, said methods comprising;

(n) retrieving from a first data-table information for a first nucleic acid, wherein said information comprises;
(oa) a first combination of letters and/or numbers uniquely identifying the nucleic acid, and
(ob) the type of genetic element comprising said nucleic acid and
(oc) a second combination of letters and/or numbers uniquely identifying a clone in which a potential member encoded by said nucleic acid was tested for interaction with at least one other potential member of a pair or complex of interacting molecules
(p) using said second combination of letters and/or numbers to retrieve from said first data-table or optionally further data-tables, information identifying additional nucleic acids encoding for said at least one other potential member in step (oc).

A preferred embodiment of said method further comprises using said second combination of letters and/or numbers in step f3) to retrieve from a second data-table further information, where said further information at least comprises the

interaction class of said clone, and optionally additional information comprising,

(q) the physical location of the clone; and
(r) predetermined experimental details pertaining to creation of said clone, including at least one of:
(ra) tissue, disease-state or cell source of the nucleic acid;
(rb) cloning details; and
(rc) membership of a library of other clones.

It is additionally preferred, that said method comprises using said information of step (o) on said first and/or of step (p) on additional nucleic acids to relate to a third data-table further characterising said first and/or additional nucleic acids, where said further characterising comprises at least one of

(s) hybridization data,
(t) oligonucleotide fingerprint data,
(u) nucleotide sequence,
(v) in-frame translation of the said nucleic acids, and
(w) tissue, disease-state or cell source gene expression data; and

optionally identifying the protein domain encoded by said first or additional nucleic acids.

Preferably also said method comprises identifying whether said potential members encoded by the nucleic acids interact, by considering said interaction class of said clone in which nucleic acids were tested for said interaction in step f3).

More preferably, said data relates to one or more of 10 to 100 potential members, yet more preferably 100 to 1000 potential members, yet more preferably, 1000 to 10000 potential members and most preferably more than 10,000 potential members.

In a preferred embodiment, said data was generated by the aforementioned method for identifying members of a pair or complex of interacting molecules.

In a further preferred embodiment, said interaction class comprises one of the following: Positive, or Negative, or False Positive.

It is further preferred, that sticky proteins are identified by consideration of the number of occurrences a given member is identified to interact with many different members in different clones of said positive interaction class.

More preferably, said first data-table forms part of a first database, and said second and third data tables form part of at least a second database.

Yet more preferably, said second database is held on a computer readable memory separate from the computer readable memory holding said first database, and said database is accessed via a data exchange network.

It is further preferred, that said second database comprises nucleic acid or protein sequence, secondary or tertiary structure, biochemical, biographical or gene expression information.

In a particularly preferred embodiment, data entry to said first, second or further data tables is controlled automatically from said first data base by access to other computer data, programs or computer controlled robots.

It is yet more preferred, that at least one workflow management system is built around particular sets of data to assist in the progress of the aforementioned method for identifying members of a pair or complex of interacting molecules.

Most preferably, said workflow management system is software to assist in the progress of the identification of members of a pair or complex of interacting molecules using the aforementioned method of hybridization of nucleic acids.

In another preferred embodiment, said data are investigated by queries of interest to an investigator.

More preferably, said queries include at least one of,

(aa) identifying the interaction or interaction pathway between a first and second member of an interaction network
(ab) identifying the interaction pathway between a first and second member of an interaction network and through at least one further member,
(ac) identifying the interaction or interaction pathway between at least two members characterised by nucleotide acid or protein sequences, secondary or tertiary structures, and
(ad) identifying interactions or interaction pathways that are different for said different tissue, disease-state or cell source.

Yet more preferably, parts of said information are stored in a controlled format to assist data query procedures.

Even more preferred is a method, wherein the results of said queries are displayed to the investigator in a graphical manner.

Yet more advantageous is the method, wherein a sub-set of data comprising data characterising nucleic acids identified as encoding members of a pair or complex of said interacting molecules is stored in a further data-table or data base.

Yet more preferably, consideration of the number of occurrences a given member is identified to interact with a second or further member is used to decide if said data characterising nucleic acids form part of said sub-set of data.

Even more preferred is the method, wherein additional information or experimental data is used to select those data to form part of said subset.

Most preferably, to speed certain data query procedures, the structure in which the data is stored in the computer readable memory is modified.

In another preferred embodiment, the data is held in relational or object oriented data bases.

The invention further relates to a data storage scheme comprising a data table that holds information on each member of an interaction, where a record in said table represents each member of an interaction, and in which members are indicated to form interactions by sharing a common name.

Preferably, in said data storage scheme said common name is a clone name or unique combination of letters and/or numbers comprising said clone name.

A computer-implemented method for handling of data gathered provides a robust and efficient solution for handling the large amount of protein-protein interaction data produced by the method of the invention. It provides the ability to communicate with and utilise different data-bases and/or other data storage systems across intra or internets, interfaces to allow querying of the data-base by an investigator and visual display of the results of the query. Relational or object orientated data-bases, with data-parsing and display programs supporting said data-base secures ease of use. By way of

example, Figure 2 displays a scheme and features for a set of data-tables suitable for managing such interaction data. The primary links between table-keys are indicated, as are the entry fields or elements to be held within each table. If desired, elements of a table may be expanded into an additional table holding further data. Likewise, certain tables may be expanded into an additional data-base to hold and manage further data. Said additional data-base may be stored on the same or on remote computers. Elements of the table can be recorded in numerical, descriptive or fixed format, whatever is most appropriate for the respective data. To provide efficient querying, where appropriate, elements are recorded in controlled vocabulary. Figure 3 displays in what part of the work process during an interaction experiment each table is most relevant and where it forms the underlying data-set from which work-flow management software for that part of the process is based.

Other computer-based methods of generating visual representations of specific interactions, partial or complete protein-protein interaction networks can be employed to automatically calculate and display the required interactions most efficiently. As is well known in the art, computer data-bases are a valuable resource for large-scale biological and molecular biological research.

In summary, a significant advantage of the method of invention over existing yeast 2H systems is the scale at which such identification of interactions and interaction members can be made. Preferably, the method of invention screens library vs. library interactions using arrayed interaction libraries. Thus, the method of invention allows, in an efficient manner, a more complete and exhaustive generation of protein-protein interaction networks than existing methods. An established and exhaustive network of protein-protein interactions is of use for many purposes as shown in Figure 1. For examples, it may be used to predict the existence of new biological

interactions or pathways, or to determine links between biological networks. Furthermore with this method, the function and localisation of previously unknown proteins can be predicted by determining their interaction partners. It also can be used to predict the response of a cell to changes in the expression of particular members of the networks. Finally, these data can be used to identify proteins or interactions between proteins within a medically relevant pathway which are suitable for therapeutic intervention, diagnosis or the treatment of a disease.

The figures show:

Figure 1

The applications of an established and exhaustive network of protein-protein interactions. The identity of positive clones and the identity of the members comprising the interactions for the entire interaction library are stored in a database. These data are used to establish a network of protein-protein interactions which can be used for a variety of purposes. For example, to predict the existence of new biological interactions or pathways, or to determine links between biological networks. Furthermore with this method, the function and localisation of previously unknown proteins can be predicted by determining their interaction partners. It also can be used to predict the response of a cell to changes in the expression of particular members of the networks. Finally, these data can be used to identify proteins within a medically relevant pathway which are suitable for therapeutic, diagnosis intervention and for the treatment of disease.

Figure 2

A scheme and features for a set of data-tables suitable for storing, managing and retrieving data from a large-scale protein-protein interaction screen. The scheme could be implemented in either relational or object-orientated data-bases. The primary links between table-keys are indicated, as

are the suggested fields or elements to be held within each table.

Figure 3

A process flow representing the experimental and informatic flow during a large-scale protein-protein interaction screen. The figure displays in which part of the experimental steps each table from a the data-base described above is most applicable. Each table forms the underlying data-set from which work-flow management software for that part of the process is based.

Figure 4

Plasmids constructed for the improved 2-hybrid system.

The plasmid maps of the pBTM118a, b and c DNA binding domain vector series and the pGAD428a, b and c activation domain vector series. Both plasmids contain the unique restriction enzyme sites for *Sal* I and *Not* I which can be used to clone a genetic fragment into the multiple cloning site. The plasmids are maintained in yeast cells by the selectable markers TRP1 and LEU2 respectively. The loss of the plasmids can be selected for by the counterselective markers CAN1 and CYH2 respectively.

Polylinkers used within the multiple cloning site to provide expression of the genetic fragment in one of the three reading frames.

Figure 5

The structure of the URA3 readout system carried by the plasmid pLUA. Important features of pLUA include the URA3 gene which is under the transcriptional control of the lexAop-GAL1 promoter, the *ADE2* selectable marker that allows yeast *ade2-* auxotrophs to grow on selective media lacking adenine and the ß-lactamase gene (bla) which confers ampicilin resistance in *E.coli*. The pLUA plasmid replicates autonomously both in yeast

using the 2μ replication origin and in *E.coli* using the ColE1 origin of replication.

Figure 6

A schematic overview of one embodiment of the method of the invention. For the parallel analysis of a network of protein-protein interactions using the method of the invention, a library of plasmid constructs that express DNA binding domain and activation domain fusion proteins is provided. These libraries may consist of specific DNA fragments or a multitude of unknown DNA fragments ligated into the improved binding domain and activating domain plasmids of the invention which contain different selectable and counterselectable markers. Both libraries are combined within yeast cells by transformation or interaction mating, and yeast strains that express potentially interacting proteins are selected on selective medium lacking histidine. The selective markers TRP1 and LEU2 maintain the plasmids in the yeast strains grown on selective media, whereas CAN1 and CYH2 specify the counter-selectable markers that select for the loss of each plasmid. HIS3 and lacZ represent selectable markers in the yeast genome, which are expressed upon activation by interacting fusion proteins. The readout system is, in the present case, both growth on medium lacking histidine and the enzymatic activity of ß-galactosidase which can be subsequently screened. A colony picking robot is used to pick the resulting yeast colonies into individual wells of 384-well microtiter plates, and the resulting plates are incubated at 30°C to allow cell growth. The interaction library held in the microtiter plates optionally may be replicated and stored. The interaction library is investigated to detect positive clones that express interacting fusion proteins and discriminate them from false positive clones using the method of the invention. Using a spotting robot, cells are transferred to replica membranes which are subsequently placed onto one of each of the selective media SD-leu-trp-his, SD-leu+CAN and SD-trp+CHX. After incubation on the selective plates, the clones which

have grown on the membranes are subjected to a ß-Gal assay and a digital image from each membrane is captured with a CCD camera which is then stored on computer. Using digital image processing and analysis clones that express interacting fusion proteins can be identified by considering the pattern of ß-Gal activity of these clones grown on the various selective media. The individual members comprising the interactions can then be identified by one or more techniques, including PCR, sequencing, hybridisation, oligofingerprinting or antibody reactions.

Figure 7

A schematic overview of one embodiment of the method of the invention. For the parallel analysis of a network of protein-protein interactions with the method of the invention, two libraries of plasmid constructs that express DNA binding domain or activation domain fusion proteins are provided. These libraries may consist of specific DNA fragments or a multitude of unknown DNA fragments ligated into binding domain and activating domain plasmids which contain the selectable markers TRP1 and LEU2, an doptionally the counterselective markers CAN1 and CYH2 respectively. The libraries are transformed into either *Mata* or *Matα* yeast strains containing the URA3 readout system and are subsequently plated onto selective media containing 5-fluoroorotic acid (5-FOA). Only those yeast cells that express fusion proteins unable to auto-activate the URA3 readout system will grow in the presence of 5-FOA. The resulting yeast strains that express only non-auto-activating proteins can then be directly used in an automated interaction mating approach to generate ordered arrays of diploid strains which can be assayed for activation of the lacZ readout system. a) Individual yeast cells that express single fusion proteins unable to activate the URA3 readout system are transferred into wells of a 384-well microtiter plate using a modified picking robot. The yeast strains held in the microtiter plates can optionally be replicated and stored. The microtiter plates contain a growth medium lacking

amino acids appropriate to maintain the corresponding plasmids in the yeast strains. The interaction matings are subsequently performed by automatically transferring a *Mata* and a *Matα* yeast strain to the same position on a Nylon membrane using automated systems as described by Lehrach *et al.* (1997). Alternatively, a pipetting or micropipetting system (Schober *et al.* 1993) can be used to transfer small volumes of individual liquid cultures of a yeast strain onto which a lawn of yeast cells derived from at least one yeast clone of the opposite mating type is sprayed or applied. Yeast strains may be applied singly or as pools of many clones. By both methods ordered arrays of yeast clones are incubated overnight at 30°C to allow interaction mating to occur. The resulting diploid cells are then analysed in a ß-Gal assay as described by Breeden & Nasmyth (1985). b) Yeast strains that grew on selective media containing 5-FOA are pooled and interaction mating between the *Mata* and *Matα* strains is made within liquid YPD medium. Those diploid yeast strains that express interacting proteins are selected by plating on selective medium lacking histidine and uracil. The selective markers TRP1 and LEU2 maintain the plasmids in yeast strains grown on selective media. HIS3, URA3 and lacZ represent reporter genes in the yeast cells, which are expressed on activation by interacting fusion proteins. The readout system is, in the present case, growth on medium lacking histidine and/or uracil and enzymatic activity of ß-galactosidase which can be screened at a later time point. A modified colony picking robot is used to pick the diploid yeast colonies into individual wells of 384-well microtiter plates containing selective medium, and the resulting plates are incubated at 30°C to allow cell growth. The interaction library optionally may be replicated and stored. Using a spotting robot, diploid cells are transferred to replica membranes which are subsequently placed onto growth medium. Replica membranes are placed on the counterselective media SD-trp+CHX or SD-leu+CAN. The resulting regular arrays of diploid yeast clones are analysed for ß-Gal activity as described by Breeden & Nasmyth

(1985). In either case a) and b), a digital image from each dried membrane is captured with a CCD camera which is then stored on computer. Using digital image processing and analysis clones that express interacting fusion proteins can be identified by considering the ß-Gal activity of these clones spotted in a defined pattern grown the membranes placed on the variious selective media. The individual members comprising the interactions can then be identified by one or more techniques, including PCR, sequencing, hybridisation, oligofingerprinting or antibody reactions.

Figure 8

Predicted interactions between fusion proteins used to create the defined interaction library. The fusion proteins enclosed with dark rounded boxes are believed to interact as shown. The LexA-HIP1 and GAL4ad-LexA fusion proteins enclosed by thin rectangular boxes have been shown to activate the LacZ readout system without the need for any interacting fusion protein. The two proteins LexA and GAL4ad, and the three fusion proteins GAL4ad-HIPCT, GAL4ad-14-3-3 and LexA-MJD (all unboxed) are believed not to interact with each other or other fusion proteins used in this example.

Figure 9

Identification of positive clones that contained interacting fusion proteins from false positive clones using the method of the invention. Three different yeast clones each containing pairs of plasmid constructs (positive control: pBTM117c-SIM1 & pGAD427-ARNT; negative control: pBTM117c & pGAD427 and false-positive control: pBTM117c-HIP1 & pGAD427) were transferred by hand to four agar plates each containing a different selective medium (SD-leu-trp, SD-leu-trp-his, SD-leu+CAN and SD-trp+CAN), and incubated for 48 hours at 30 °C. The yeast colonies were subsequently transferred to a Nylon membrane and assayed for β-gal activity by the method of Breeden and Nasmyth (1985).

Figure 10

Digital images of the ß-gal assays made from the replica Nylon membranes containing the defined interaction library obtained from the selective media (a) SD-leu-trp-his, (b) SD-trp+CHX and (c) SD-leu+CAN. In each case, The left hand side of each membrane contains control clones and clones from the defined interaction library, and the right hand side contains only clones from the defined interaction library. The two regions marked on the first membrane represent those clones magnified in Figure 11. The overall size of each membrane is 22 x 8 cm and contains 6912 spot locations at a spotting pitch of 1.4 mm.

Figure 11

Magnification of clones from the interaction library taken from the same region of three membranes obtained from the selective media SD-leu-trp-his, SD-trp+CHX and SD-leu+CAN assayed for β-gal activity:

Clones imaged from a region of the right hand side of the membrane containing the defined interaction library. Clones from the defined interaction library that express interacting proteins are ringed and correspond to the microtiter plate addresses 06L22 and 08N24.

Clones imaged from a region of the left hand side of the same membranes containing control clones and clones from the interaction library, where clones around each ink guide-spot are arranged as shown and correspond to: 00 Ink guide spot; 01 False positive control clone that expresses the fusion protein GAL4ad-LexA; 02 False positive clone expressing the fusion protein LexA-HIP1; 03 Positive control clone expressing the interacting fusion proteins LexA-SIM1 & GAL4ad-ARNT; 04 Clone from the defined interaction library. The positive control clone (spot position 03) is ringed.

Figure 12

A subset of the list of clones identified by computer query of data produced by automated image analysis and quantification of the ß-galactosidase activity. Each record represents the ß-galactosidase activity for a given clone grown on three selective media. This program queried the data to identify all clones from the interaction library that had activated the reporter gene (score > 0) when grown on minimal medium lacking, leucine, trptophan, and histidine (SD-leu-trp-his), yet had not on either of the counterselective media (score on both media equal to 0).

Two positive clones 06L22 and 08N24 characterised by hybridisation are present within the computer file.

Figure 13

Characterisation by hybridisation of the genetic fragments carried by the clones 06L22 and 08N24. A 1.3 kb, SIM1 and a 1.4 kb ARNT DNA fragment were used as nucleic acid probes for hybridisation to high-density spotted membranes containing DNA from the defined interaction library. These clones were characterised as containing SIM1 and ARNT genetic fragments by hybridisation. The images are of the same region of the membranes as those shown in Figure 11 a. The spot positions of the clones 06L22 and 08N24 are ringed.

Figure 14

Identification of the SIM1 and ARNT DNA fragments from the yeast two hybrid plasmid carried by the clone 06L22 by duplex PCR. Plasmid DNA was isolated from a liquid culture of the clone 06L22 by a QiaPrep (Hilden) procedure and the inserts contained within the plasmids were amplified by PCR using the primer pairs, 5'-TCG TAG ATC TTC GTC AGC AG-3' & 5'-GGA ATT AGC TTG GCT GCA GC-3' for the plasmid pBTM117c and 5'-CGA TGA TGA AGA TAC CCC AC-3' & 5'-GCA CAG TTG AAG TGA ACT TGC-3' for pGAD427. Lane 1 contains a Lamda DNA digestion with *Bst*EII as size marker; Lane 2 contains the duplex PCR reaction from

plasmids isolated from clone 06L22; Lanes 3 and contain control PCR amplifications from the plasmids pBTM117c-SIM1 and pGAD427-ARNT respectively.

Figure 15

Readout system ativation for clones in a regular grid pattern from an interaction library. 23 384-well microtiter plates of the sea urchin interaction library were spotted in a `3x3 duplicate' regular grid pattern around an ink guide-spot on a 222 x 222 mm porous membrane (Hybond N+, Amersham, UK) using a spotting robot. The membrane was incubated in SD-leu-trp-his medium for 3 days, assayed for lacZ expression using the ß-gal assay as described by Breeden & Nasmyth (1985) and air dried overnight. A digital image was captured using a standard A3 computer scanner.

Figure 16

Hybridisation of a gene fragment (Probe A) encoding for Protein A to an array of DNA from an interaction library. The probe was labelled radioactively by standard protocols, and hybridisation-positive clones from the interaction library are identified by the automated image analysis system. The position of clone 5K20, from which the gene fragment was isolated, is indicted. Other hybridisation-positive clones also carry this gene-fragment, and by recovery of interacting members from these clones, a protein-protein interaction pathway for Protein A can be uncovered.

Figure 17

A graphical representation of the hybridisation-positive clones generated by hybridisation of Probe A to a DNA array representing the interaction library.

Figure 18

A graphical representation of hybridisation- and interaction-positive clones generated by a subsequent hybridisation with probe B (isolated from the clone marked in a grey box). Also

shown, are the positions of the hybridisation-positive clones from probe A. Interaction-positive clones that carry both gene fragments are identified as hybridising with both probes.

Figure 19

A graphical representation of hybridisation- and interaction-positive clones generated by a further hybridisation with probe C isolated from the clone 6D18 (marked by a grey box and "B/C"). Also shown are the hybridisation signals for probes A and B. By considering common hybridisation signals for interaction-positive clones and subsequent DNA sequencing of the inserts carried by these clones, protein-protein interactions can be uncovered. The figure also shows an interaction pathway uncovered between Proteins A, B an C based on these data.

Figure 20

Automated visual differentiation of yeast cells expressing single fusion proteins able to activate the LacZ readout system. A defined library of L40ccu yeast clones expressing different fusion proteins cloned in the plasmid pBTM117c was plated onto minimal medium lacking tryptophan, buffered to pH 7.0 with potassium phosphate and containing 2 ug/ml of X-Gal (SD-trp/XGAL). White colonies that have not autoctivated the LacZ reporter gene are automatically recognised and marked with a red horizontal cross. A colony that has turned blue due to expression of a single fusion protein able to auto-activate the LacZ reporter gene is automatically recognised due to its darker colour and the presence of a `hole'. An arrow indicates this colony. All colonies unsuitable for further analysis and picking (including those too small or touching colonies) are automatically recognised and marked with a blue diagonal cross.

Figure 21

Results of automated interaction mating to identify diploid yeast strains that express interacting fusion proteins. a)

Progeny of the yeast strains x1a and x2a were spotted at positions 1 and 2 on a nylon membrane using a spotting robot such as described by Lehrach et al. (1997). The yeast strains y1α and y2α of the opposite mating type were subsequently spotted on positions 1 and 2 which already contained cells from the strains x1a and x2a. To assist in recognition of the duplicate spotting pattern, ink was spotted in position 2 directly to the right of the spotted yeast clones. b) The membrane was transferred to a YPD agar plate and was incubated at 30° C overnight to allow interaction mating to occur. c) Diploid yeast cells that had grown on the membrane were subsequently analysed for ß-galactosidase activity using the method of Breeden & Nasmyth (1985).

Figure 22

The two vectors constructed to provide further genetic features to enable the method of invention within a prokaryotic two-hybrid system. The vectors are based on the pBAD series of vectors which provide tight inductive-control of expression of cloned genes using the promoter from the arabinose operon (Guzman et al., 1995 J. Bact. 177: 4141-4130), and can be maintained in the same *E.coli* cell by virtue of compatible origins of replication.

The plasmid pBAD18-αRNAP expresses under the control of the arabiose promoter, fusion proteins between the α amino terminal domain (NTD) of the α-subunit of RNA polymerase and DNA fragments cloned into the multiple cloning site. The presence of this plasmid in kanamycin sensitive cells can be selected by plating on growth medium supplemented with kanamycin, or for its absence by the counterselective *rpsL* allele by plating on media supplemented with streptomycin (Murphy et al. 1995).

The plasmid pBAD30-cI expresses under the control of the arabinose promoter, fusion proteins between the λcI protein and DNA fragments cloned into the multiple cloning site. The

presence of this plasmid in amplicillin sensitive cells can be selected by plating on growth medium supplemented with amplicillin, or for its absence by the counterselective *lacY* gene by plating on media supplemented with 2-nitrophenyl-ß-D-thiogalactosidase (tONPG) (Murphy et al. 1995). Additionally, the oriT sequence enables unidirectional genetic exchange of the pBAD30-cI plasmid and its derivatives from E.coli cells containing the F' fertility factor to F^- strains lacking the fertility factor.

Examples

Example 1: Construction of vectors yeast strains and readout system for an improved yeast two-hybrid system

1.1 Construction of vectors

The plasmids constructed for an improved yeast two-hybrid system pBTM118 a, b and c and pGAD428 a, b and c are shown in Figure 4. Both sets of vectors can be used for the construction of hybrid (fusion) proteins. The vectors contain the unique restriction sites *Sal* I and *Not* I located in the multiple cloning site (MCS) region at the 3´- end of the open reading frame for either the *lexA* coding sequence or the GAL4ad sequence Figure 4b).

With both sets of plasmids fusion proteins are expressed at high levels in yeast host cells from the constitutive *ADH1* promoter (P) and the transcription is terminated at the *ADH1* transcription termination signal (T). The two-hybrid plasmids shown in Figure 4a are shuttle vectors that replicate autonomously in both *E. coli* and *S. cerevisiae*.

The three plasmids pBTM118 a, b and c are used to generate fusions of the LexA protein (amino acids 1-220) and a protein of interest cloned into the MCS in the correct orientation and

reading frame. The plasmids pBTM118 a, b and c are derived from pBTM117c (Wanker et al., 1997) by insertion of the adapters shown in Table 1 into the restriction sites *Sal* I and *Not* I to generate the improved vectors with three different reading frames.

The plasmids pBTM118 a, b and c carry the wild type yeast *CAN1* gene for counterselection, which confers sensitivity to canavanine in transformed yeast cells (Hoffmann, 1985). The plasmids also contain the selectable marker *TRP1*, that allows yeast trp1-auxotrophs to grow on selective synthetic medium without tryptophan, and the selectable marker *bla* which confers ampicillin resistance in *E. coli*.

The plasmids pGAD428 a, b and c are used to generate fusion proteins that contain the GAL4 activation domain (amino acids 768-881) operatively linked to a protein of interest. The plasmids pGAD428 a, b and c carry the wild type yeast *CYH2* gene, which confers sensitivity to cycloheximide in transformed cells (Kaeufer et al., 1983), the selectable marker *LEU2*, that allows yeast leu2-auxotrophs to grow on selective synthetic medium without leucine, and the bacterial marker *aphA* (Pansegrau et al., 1987) which confers kanamycin resistance in *E. coli*. The plasmids pGAD428a, b and c were created from pGAD427 by ligation of the adapters shown in Table 1 into the MCS to construct the improved vectors with three different reading frames.

For the construction of pGAD427 a 1.2 kb *Dde* I fragment containing the *aphA* gene was isolated from pFG101u (Pansegrau et al., 1987) and was subcloned into the *Pvu* I site of the pGAD426 using the oligonucleotide adapters 5'- GTCGCGATC-3' and 5'-TAAGATCGCGACAT-3'. The plasmid pGAD426 was generated by insertion of a 1.2 kb *Eco* RV *CYH2* gene fragment, which was isolated from the pAS2-1 (Clonetech) into the *Pvu* II site of pGAD425 (Han and Collicelli, 1995).

1.2 Construction of yeast strains

To allow for the improved yeast two-hybrid system, three *Saccharomyces cerevisiae* strains L40cc, L40ccu and L40ccuα were created. The *S. cerevisiae*. L40cc was created by site specific knock-out of the CYH2 and CAN1 genes of L40 (Hollenberg et al., Mol. Cell. Biol. 15: 3813-3822), and L40ccu created by site specific knock-out of the URA3 gene of L40cc (Current Protocols in Molecular Biology, Eds. Ausubel et al. John Wiley & Sons: 1992) The strain L40ccuα was created by conducting a mating-type switch of the strain L40ccu by standard procedures (Ray BL, White CI, Haber JE (1991)). The genotype of the L40cc strain is: *Mata his3Δ200 trp1-901 leu2-3,112 ade2 LYS2::(lexAop)$_4$-HIS3 URA3::(lexAop)$_8$-lacZ GAL4 can1 cyh2*, The genotype of the L40ccu strain is: *Mata his3Δ200 trp1-901 leu2-3,112 ade2 LYS2::(lexAop)$_4$-HIS3 ura3::(lexAop)$_8$-lacZ GAL4 can1 cyh2*, and that of L40ccuα is *Matα his3Δ200 trp1-901 leu2-3,112 ade2 LYS2::(lexAop)$_4$-HIS3 ura3::(lexAop)$_8$-lacZ GAL4 can1 cyh2*.

1.3 Readout system

Figure 5 shows the *URA3* readout system carried by the plasmid pLUA. This *URA3* readout system under the control of a bacterial LexAop upstream activation sequence (UAS) can be used within the yeast 2-hybrid system both as a counter selective reporter gene and as a positive selection reporter gene to eliminate false positive clones. The plasmid contains the features of the UAS$_{lexAop}$-URA3 readout system, the selectable marker *ADE2* that allows yeast ade2-auxotrophs to grow on selective media without adenine and the *bla* gene which confers amplicillin resistance in *E.coli*. The plasmid pLUA is a shuttle vector that replicates autonomously in *E. coli* and yeast.

For the construction of pLUA a 1.5 kb *Sac* I/*Cla* I UAS$_{lexAop}$-URA3 fragment was isolated from pBS-lexURA and ligated

together with a 2.4 kb *Sac* I/*Cla* I *ADE2* fragment into *Cla* I digested pGAD425Δ. pBS-lexURA was generated by ligating *URA3* fragment together with a UAS_{lexAop} fragment into pBluescript SK+. The *URA3* and UAS_{lexAop} fragments were obtained by PCR using genomic DNA from *S. cerevisiae* strain L40c using standard procedures and anchor primers which gave rise to complementary overhangs between the two consecutive fragments which were subsequently anealed to generate the chimeric sequence (see, for example, Current Protocolls in Molecular Biology, Eds. Ausubel et al. John Wiley & Sons: 1992). The *ADE2* gene was isolated by PCR using genomic DNA from SEY6210α. pGAD425Δ was generated by deleting of an 1.2 kb *Sph* I fragment from pGAD425 (Han and Colicelli, 1995) and religation of the vector.

1.4 Generation of a defined interaction library

To determine if the invention could be used in an improved two-hybrid system for yeast, as shown in Figure 6 or Figure 7, a defined interaction library of plamids that express various LexA and GAL4ad fusion proteins of interest was constructed using the vectors and strains described in sections 1.1 and 1.2. The orientation of the inserted fragments was determined by restriction analysis and the reading frame was checked by sequencing. The generated constructs and the original plasmids described above are listed in Table 2. The construction of pBTM117c-HD1.6, -HD3.6 and -SIM1 was described elsewhere (Wanker et al., 1997; Probst et al., 1997). pBTM117c-HIP1 and pGAD427-HIP1 were obtained by ligation of a 1.2 kb *Sal* I HIP1 fragment isolated from pGAD-HIP1 (Wanker et al., 1997) into pBTM117c and pGAD427, respectively. pBTM117c-MJD was created by inserting a 1.1 kb *Sal* I/*Not* I MJD1 fragment (Kawagushi et al., 1994) into pBTM117c, and pGAD427-14-3-3 was generated by inserting a 1.0 kb EcoRI/NotI fragment of pGAD10-14-3-3 into pGAD427. For the construction of pGAD427-HIPCT, a 0.5 kb *Eco* RI HIP1 fragment isolated from pGAD-HIPCT (Wanker et al., 1997) was ligated into pGAD427. pGAD427-lexA and pGAD427-ARNT

were generated by insertion of a 1.2 kb *Sal* I/*Not* digested lexA PCR fragment and a 1.4 kb *Sal* I/*Not* I ARNT fragment into pGAD427 respectively.

It was shown that the fusion proteins LexA-SIM1 and GAL4ad-ARNT specifically interact with each other in the yeast two-hybrid system (Probst et al., 1997), because when both hybrids were coexpressed in *Saccharomyces cerevisiae* containing two integrated reporter constructs, the yeast *HIS3* gene and the bacterial *lacZ* gene, which both contained binding sites for the LexA protein in the promoter region, the interaction between these two fusion proteins led to the transcription of the reporter genes. The fusion proteins by themselves were not able to activate transcription because GAL4ad-ARNT lacks a DNA binding domain and LexA-SIM1 an activation domain (Probst et al., 1997). In contrast it was shown recently that the fusion proteins LexA-HIP1 and GAL4ad-LexA are capable of activating the *HIS3* and *lacZ* reporter genes without interacting with a specific GAL4ad or LexA fusion protein respectively. Thus, the yeast clones expressing the LexA-HIP1 protein have to be designated as false positives, because false positives are defined here as clones where a GAL4ad fusion protein or a LexA fusion protein alone without the respective partner protein activates the transcription of the reporter genes without the need for any interacting partner protein.

The predicted protein-protein interactions of these fusion proteins are shown in Figure 8. It was shown that the fusion proteins LexA-SIM1 & GAL4ad-ARNT, LexA-HD1.6 & GAL4ad-HIP1 and LexA-HD3.6 & GAL4ad-HIP1 specifically interact with each other in the yeast two-hybrid system because they only activate the reporter genes *HIS3* and *lacZ* when both proteins are present in one cell (Probst et al. 1997; Wanker et al. 1997). In contrast, it was demonstrated that the LexA-HIP1 and GAL4ad-LexA fusion proteins are capable of activating the reporter genes without the need for any interacting fusion protein. The proteins LexA and GAL4ad and the fusion proteins LexA-MJD and

GAL4ad-14-3-3 which are also present in the defined interaction library are unable to activate the reporter genes either alone or when present in the same cell with any other fusion proteins comprising the library.

Example 2: Detection of clones expressing known interacting proteins from false positives using the improved two-hybrid system

Pairs of the yeast two-hybrid plasmids pBTM117cSIM1 & pGAD427-ARNT, pBTM117c & pGAD427 and pBTM117c-HIP1 & pGAD427 were transformed into the yeast strain L40cc, and Trp+Leu+ transformants that contained at least one of each of the two plasmids were selected on SD-leu-trp plates. Two transformants from each transformation were investigated for the presence of protein-protein interactions by testing the ability of the yeast cells to grow on SD-leu-trp, SD-leu-trp-his, SD-leu+CAN and SD-trp+CHX plates and by the ß-galactosidase membrane assay (Breeden and Nasmyth, 1985). Figure 9 shows that the yeast strains cells harboring both the plasmids pBTM117c-SIM1 & GAD427-ARNT or pBTM117c-HIP1 & pGAD427 grow on SD-leu-trp-his plates and turned blue after incubation in X-Gal solution, indicating that the *HIS3* and *lacZ* reporter genes are activated in these strains. In comparison, the yeast strain harboring both the negative control plasmids pBTM117c & pGAD427 was not able to grow on this medium and also showed no lacZ activity. After selection of the yeast strains harboring the different combinations of the two-hybrid plasmids on SD-leu+CAN and SD-trp+CHX the resulting strains were also analyzed by the ß-galactosidase assay. After incubating the membrane containing all three strains on SD-trp+CHX medium only progeny of the yeast strain that originally harbored both the plasmids pBTM117c-HIP1 & pGAD427 yet which had lost the pGAD427 plasmid through counterselection turned blue after incubating in X-Gal solution. This result indicates that this clone is a false positive, because although showing a lacZ+ phenotype when grown on SD-leu-trp-his medium, the LexA-HIP1 fusion protein

was also capable of activating the HIS3 and lacZ genes on SD-trp+CAN medium without the need for any interacting fusion protein. In comparison, the yeast strain harboring both the plasmids pBTM117c-SIM1 & pGAD427-ARNT is a positive clone that expresses interacting LexA and GAL4ad fusion proteins, because both the LexA and the Gal4ad fusion proteins are necessary for the activation of the reporter genes. If either of the plasmids pBTM117c-SIM1 or pGAD427-ARNT is lost from the strain by counterselection on SD-trp+CHX or SD-leu+CAN, respectively, the resulting cells are no longer able to activate the lacZ reporter gene and do not turn blue after incubation in X-Gal solution. With the membranes from the SD-leu+CAN plate false positive clones expressing an auto-activating GAL4ad-LexA fusion protein were also detected by the ß-galactosidase assay.

Example 3: Generation of regular grid patterns of host cells expressing potentially interacting fusion proteins

3.1 Generation of a regular grid pattern of clones from an interaction library in microtiter plates using automation

To generate the well defined interaction library, the constructs for the expression of the fusion proteins shown in Figure 8 were pooled and 3 μg of the mixture was co-transformed into yeast strain L40cc by the method of Schiestel & Gietz (1989). The yeast cells co-transformed with the constructs described in Table 2 were plated onto large 24 x 24 cm agar trays (Genetix, UK) containing minimal medium lacking tryptophan leucine and histidine (SD-leu-trp-his). The agar trays were poured using an agar-autoclave and pump (Integra, Switzerland) to minimise tray-to-tray variation in agar colour and depth. To maximise the efficiency of automated picking, the transformation mixture were plated such that between 200 and 2000 colonies per agar tray were obtained after incubation at 30°C for 4 to 7 days.

Suitable changes to the hardware and software of a standard picking robot designed for the picking of *E. coli* cells as described by Lehrach et al. (1997) were made to account for the specific requirements of yeast cells. The illumination of agar-trays containing plated colonies was changed from the dark-field sub-illumination to dark-field top-illumination to differentiate yeast colonies from the lawn of non-transformed cells. The existing vision guided motion system (Krishnaswamy & Agapakis 1997) was modified to allow for a larger range of `blob' size when selecting yeast colonies to pick from the blob-feature-table returned by connectivity algorithms when applied to a digital image of the agar tray containing colonies. The clone inoculation routine was re-programmed to ensure that cell material which had dried on the picking pins during the picking routine was initially re-hydrated by 10 seconds of immersion in the wells of a microtiter plate before vigorous pin-motion within the well. This robotic procedure ensured that sufficient cell material was inoculated from each picking pin into an individual well of a microtiter plate. The picking pins were sterilised after inoculation to allow the picking cycle to be repeated by programming the robot to brush the picking pins in a 0.3% (v/v) solution of hydrogen peroxide, followed by a 70% ethanol rinse from a second wash-bath and finally drying by use of a heat-gun to evaporate any remaining ethanol from the pins. Furthermore, an algorithm to automatically correct for height variation in the agar was incorporated by referencing the surface height of the agar in three corners and from these points automatically estimating the surface plane of the agar. The robot was further programmed to automatically adjust both the imaging and picking heights according to the agar surface height such that when a pin was extended into a colony, it removed cells only from the top surface of the colony and did not penetrate the whole colony into the growth medium. Finally, we incorporated additional selection criteria that would reliably sort between blue and white colonies. Although the robot provided a method to select only those `blobs' (colonies) within a range of

average grey scales (eg, > 80 for white colonies), this proved unreliable since the actual value of average grey scale required to make a correct discrimination varied across the agar-tray due to slight variation in intensity of the illumination. Therefore, a new method was implemented that automatically modified this discrimination value based on the average illumination of a region of the agar-tray as measured by the camera on a frame-to-frame basis. Often, a `blue' colony that activated the readout system was not uniformly blue across the its whole area, but only the centre would be blue and the surrounding cell material was white. In such cases, the connectivity algorithms would detect two `blobs' - one (the blue centre) lying directly on the other (the white surrounding) and although the former would be ignored since it was blue, the latter would be selected as its average grey-scale was greater than the discrimination value. Such cases were successfully selected against by ignoring any colonies that had `holes' using a `number of holes' function of the image analysis program, which flags those blobs which have a second blob within their boundary.

Using these modifications to a laboratory picking robot, individual yeast colonies were automatically picked from the agar-trays into individual wells of a sterile 384-well microtiter plate (Genetix, UK) containing sterile liquid minimal medium lacking leucine and trptophan (SD-leu-trp) and containing 7% (v/v) glycerol. The resulting microtiter plates were incubated at 30°C for 36 hours, the settled colonies were dispersed by vigorous mixing using a 384-well plastic replicating tool (Genetix, UK) and then incubated for a further 2 to 4 days. A picking success of over 90% wells containing a growing yeast culture was achieved. After growth of yeast strains within the microtiter plates, each plate was labelled with a unique number and barcode. Each plate was also replicated to create two additional copies using a sterile 384-pin plastic replicator (Genetix, UK) to transfer a small amount of cell material from each well into pre-labelled 384-

well microtiter plates and pre-filled with SD-leu-trp-his/7% glycerol liquid medium. The replicated plates were incubated at 30 °C for 3 days with a cell dispersal step after 36 hours, subsequently frozen and stored at -70°C together with the original picked microtiter plates of the interaction library.

In this manner, a regular grid pattern of yeast cells expressing potentially interacting yeast clones was generated using a robotic and automated picking system. 384-well microtiter plates have a well every 4.5 mm in a 16 by 24 well arrangement. Therefore, for each 384-well microtiter plate a regular grid pattern at a density greater that 4 clones per square centimetre was automatically created.

3.2 Creation of regular grid patters of increased density

To generate arrays with higher densities, a computer-controlled 96-well pipetting system (Opal-Jena) with automatic plate-stacking, tip washing, liquid waste and accurate x-y positioning of the microtiter plate currently accessed by the tips was employed. The yeast two hybrid cells that had settled in the bottom of the wells of the arrayed interaction library as described above were re-suspended, and a stack of these 384-well plates were placed into the input stacker of the pipetting system. The system was programmed to take a single 384-well microtiter plate containing the arrayed yeast two-hybrid clones and parallel aspirate 10 μl of culture medium and cells into each of the 96 pipette tips from 96 wells of the 384-well plate. The inter-tip spacing of the 96 tips was 9mm and the wells of the 384-well microtiter plate were 4.5 mm so that cells were removed from only every other well along each dimension of the 384-well plate. 8 μl of the 96 aspirated samples contained in the tips were then pipetted in parallel into one set of wells of a sterile 1536-well microtiter plate (Greiner, Germany). Since the inter-well spacing of this 1536-well microtiter plate is 2.25 mm, yeast cells were deposited into only 1 every 4 wells along each dimension of the 1536-

well plate. The remaining 2 μl of culture medium and cells was aspirated to waste before sterilising each 96 tips in parallel. Sterilisation was conducted by twice aspirating and washing to waste 50 μl of 0.3% (v/v) hydrogen peroxide stored in a first replenishable wash-bath on the system, and then aspirating and washing to waste 50 μl sterile distilled water stored in a second replenishable wash-bath.

This plate-to-plate pipetting cycle was repeated 3 further times, each time aspirating a different set of 96-clones from the 384-well array of yeast 2-hybrid clones into a different set of 96-wells in the 1536-well microtiter plate by moving the microtiter plates relative to the 96-tips using the accurate x-y positioning of the system. When all clones of the first 384-well microtiter plate had been sampled and arrayed into the 1536-well plate, the first 384-well microtiter plate was automatically exchanged for the next 384-well microtiter plate, and the yeast 2-hybrid clones arrayed in this second 384-well plate were similarly arrayed into the 1536-well plate. When the yeast 2-hybrid clones contained within four 384-well microtiter plates had been automatically arrayed in the first 1536-well plate, filling all wells, the 1536-well plate was automatically exchanged for a second sterile 1536-well plate stored in the second stacking unit of the pipetting system. The whole process was repeated until all yeast 2-hybrid clones of the interaction library had been automatically transferred form 384-well to 1536-well microtiter plates.

In this manner, a regular grid pattern of yeast cells expressing potentially interacting yeast clones using a computer-controlled pipetting system was generated. 1536-well microtiter plates have a well every 2.25 mm in a 32 by 48 well arrangement. Therefore, for each 1536-well microtiter plate we automatically created a regular grid pattern at a density greater than 19 clones per square centimetre.

are flanked by 5´and 3´LTRs, respectively. pREV-TRE is available from Clonetech Inc.. cDNA libraries are subcloned into the polylinker of pREV-TRE.
The above described reporter cell lines are separately infected with either pREV-TRE-Gal4- or pREV-TRE-VP16-derived retroviral particles. A polyclonal pool of new stable cell lines is selected in both cases using the resistance selection marker gene hyg^r. Transient expression of fusion proteins from pREV-TRE plasmids has to be induced by withdrawal (Tet-Off) or addition (Tet-On) of Tet in order to allow for double preselection and elimination of false positives as described above.

11.7 Cell Fusion and Selection for Cells Expressing Interacting Proteins

The remaining polyclonal pools of cell lines are then subjected to cell fusion as described above. The HygB concentration in the culture medium is increased to minimize a possible loss of either one component of the pairs of fusion protein cDNA sequences present in all fused cells. For the detection of positive clones, i.e. cells expressing a pair of interacting proteins (as detailed above), expression of fusion proteins has to be induced by addition or withdrawal of Tet.

References:

Allen, T.B., Walberg, M.W., Edwards, M.C., Elledge, S.J. (1995) Finding prospective partners in the library: the two-hybrid system and phage display find a match. TIBS, 20: 511-516

Anderson M.T., Tjioe I.M., Lorincz M.C. Parks D.R., Herzenberg L.A., Nolan G.P., Herzenberg L..A. (1996) Simultaneous fluorescence-activated cell sorter analysis of two distinct transcriptional elements within asingle cell using engineered green fluorescent proteins. Proc. Natl. Acad.Sci. U S A: 93: 8508-11

Barillo, E., Lacroix, B & Cohen, D (1991) Theoretical analysis of library screening using a N-dimensional pooling strategy. NAR 25: 6241-6247.

Bartel, P., Chien, C.-T., Sternglanz, R., Fields, S. (1993) Elimination of false positives that arise in using the two-hybrid system. Biotechniques 14: 920-924.

Bartel, P.L., Roecklein, J.A., SenGupta, D., Fields, S. (1996) A protein linkage map of Escherichia coli bacteriophage T7. Nat. Genet. 12: 72-77

Bendixen, C., Gangloff, S., Rothstein, R. (1994) A yeast mating-selection scheme for detection of protein-protein interactions. Nucl. Acids Res., 22: 1778-1779.

Benton, D. (1996) Bioinformatics--principles and potential of a new multidisciplinary tool. Trends in Biotechnology 14: 261-272.

Breeden, L. and K. Nasmyth, K. (1985). Regulation of yeast HO gene. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 643-650.

Boeke, J.D., LaCroute, F. and Fink, G.R. (1984). A positive selection for mutants lacing orotidine-5'-phosphate decarboxylase activitiy in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. 197: 345-346

Cubbitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. and Tsien, R.Y. (1995). Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20**: 448-455.

Davidson, 1986. Gene Activity in Early Development, Third Edition, Academic press, Orlando Florida.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**: 457-460.

Dove, S., Joung, J.K., Hochschild, A. (1997) Activation of prokaryotic transcription through arbitrary protein-protein contacts. Natur, 386:627-630

Dramanac, R., Labat, I., Brukner, I., Crkvenjakov, R. (1989) Sequencing of megabase plus DNA by hybridisation: theory of the method. Genomics, 4: 114-128.

Fearon, E., Finkel, T., Gillison, M.L., Kennedy, S.P., Casella, J.F., Tomaselli, G.F., Morrow, J.S., Van Dang, C. (1992) Karyoplasmic interaction selection strategy: A general strategy to detect protein-protein interactions in mammalian cells. Proc. Natl. Acad. Sci. USA, 89: 7958-7962.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specifity activity. *Anal. Biochem.* **132**: 6-13

Fields, S.and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.

Fleischmann, R.D., Adams, M.D., (1995) Whole genome random sequencing and assembly of *Haemophiilus influenza* Rd. Science 269: 496-512.

Gietz, D., St. Jean, A., Woods, R.A., Schiestl, R.H. (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Research 20: 1425.

Gress, T.M., Müller-Pillasche, F., Geng, M., Zimmerhack, F., Zehetner, G., Friess, H., Büchler, M., Adler, G., Lehrach, H. (1996) A cancer-specific expression profile. Oncogene 13: 1819-1830.

Han, L. and Colicelli, J. (1995). A human protein selected for interference with Ras function interacts directly with Ras and competes with Raf1. *Mol. Cell. Biol.* **15**: 1318-1323.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., Elledge, S.J. (1993) The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 Cyclin-dependent kinases. Cell, 75: 805-816.

Hoffmann, W. (1985). Molecular characterisation of the CAN1 locus in Saccharomyces cerevisiae. *J. Biol. Chem.* **260**: 11831-11837.

Hoheisel, J.D., Lennon, G.G., Zehetner, G. & Lehrach, H 1991. Use of reference libraries of *Drosophila melanogaster* for relational data analysis; a step towards mapping and sequencing of the genome. J. Mol. Biol. 20: 903-914.

Hurd, D., Fallon, R.A., White, M., Jones, N. (1997) Improvements relating to assay systems. WO 97/23609

Johnson, R.F., Pickett, S.C., Barker, D.L. (1990) Autoradiography using storage phosphor technology. Electrophoresis 11: 355-360.

Kaeufer, N.F., Fried, H.M., Schwindinger, W.F., Jasin, M. and Warner, J.R. (1983). Cycloheximide resistance in yeast: the gene and its protein. *Nucleic Acids Res.* **11**: 3123-3135.

Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., Narumiya, S. and Kakizuka, A. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.* **8**: 221-228.

Kietzmann, M., Kalkum, M., Maier, E., Bancroft., David, Eickhoff, H., Ivanov, I., Przewieslik, T., Horn, M. & Lehrach, H. (1997) Pizo-ink-jet based pipetting-system for high density gridding and nanowell filling. Poster presentation at: *Automation in mapping and DNA sequencing*. EMBL Heidelberg, March 16-19th 1997.

Larin, Z. and Lehrach, H. (1990). Yeast artificial chromosomes: an alternative approach to the molecular analysis of mouse developmental mutations. *Genet. Res.* **56**: 203-208.

Lehrach, H., Bancroft, D. and Maier, E. (1997). Robotics, computing, and biology: An interdisciplinary approach to the analysis of complex genomes. *Interdisp. Science Rev.* **22**: 37-43.

Le Douarin, B., Pierrat, B., vom Baur, E., Chambon, P,. Losson, R. (1995) A new version of the two-hybrid assay for detection of protein-protein interactions. Nucl. Acids Res., 23: 876-878.

Lennon, G.G., Lehrach, H. (1991) Hybridisation analysis of arrayed cDNA libraries. Trends in Genetics 7: 375-388.

Li, M. (1997) Compounds and related methods for modulating potassium ion channels and assays for such compounds. WO 97/31112.

Li, J.J. and Herskowitz, I. (1993). Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* **262**:1870-1874.

Link A.J., Phillips D. & Church G.M. (1997) Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterisation. J Bacteriol 179:6228-6237

Liu, J. Stanton, VP, Fujiwara, TM, Wang, JX, Rezonzew, R. Crumley, MJ, Morgan, K, Gros, P., Housman, D. & Schurr, E (1995) large-scale cloning of human chromosome 2-specific yeast artificial chromosomes (YACs) using an interspersed repetitive sequences (IRS)-PCR approach. Genomics 26: 178-191.

Lockhart, D. J., Dong, H. Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E. L. (1996) Expression monitoring by hybridisation to high-density oligonucleotide arrays. nature Biotechnology 12: 1675-1680.

Maier, E., Maier-Ewert, S., Bancroft, D., Lehrach, H. (1997) Automated array technologies for gene expression profiling. Drug Discovery Today, 2: 315-324.

Meier-Ewert, S., Maier, E., Ahmadi, A., Curtis, J & Lehrach, H. (1993) An automated approach to generating expressed sequence catalogues. Nature 361: 375-376.

Murphy, C.K., Stewart, E.J. & Beckwith J. (1995) A double counter-selective system for the study of null allelels of essential genes in *Eschericia coli*. Gene 155: 1-7.

Nandabalan, K., Rothberg, J.M., Yang, M., Knight, J.R., Kalbfleisch, T. (1997) Identification and comparison of protein-protein interactions and inhibitors thereof, WO 97/47763

Pansegrau, W., Miele, L., Lurz, R. and Lanka E. (1987). Nucleotide sequence of the kanamycin resistance determinant of plasmid RP4: homology to other aminoglycoside 3'-phosphotransferases. *Plasmid* **18**: 193-204.

Probst, M.R., Fan, C.M., Tessier-Lavigne, M. and Hankinson, O. (1997). Two murine homologs of the Drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. *J. Biol. Chem.* **272**: 4451-4457.

Putz, U., Skehel, P. and Kuhl, D. (1996). A tri-hybrid system for the analysis and detection of RNA--protein interactions. *Nucleic Acids Res.* **24**: 4838-4840 .

Ray BL, White CI, Haber JE (1991) Heteroduplex formation and mismatch repair of the "stuck" mutation during mating-type switching inSaccharomyces cerevisiae. Mol Cell Biol 11:5372-80

Ross, M.T., Hoheisel, J.D., Monaco, A.P., Larin, Z., Zehetner, G., & Lehrarch, H. (1992) High density gridded YAC filters: their potential as genome mapping tool. In: Anand, R (Ed.) Techniques for the analysis of complex genomes. Academic Press, New York, pp. 137-154.

Schiestl, R.H. and Gietz, R.D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**: 339-346.

Schober, A., Guenther, R. Schwienhorst, A., Doering, M. and Lindemann, B. F. (1993). Accurate high-speed liquid handling of very small biological samples. Biotechniques 15: 324-329.

SenGupta DJ., Zhang B., Kreamer B., Pochart P., Fields S., Wickens M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. *PNAS*. 93:8496-501

Shalon, D., Smith, S. J. and Brown, P. O. (1996). A DNA microarray system for analysing complex DNA samples using two-colour fluorescent probe hybridisation. Geneome Research 6, 639-645.

Sherman, F., Fink, G.R. & Hicks, J.B. Methods in Yeast Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Stanton, R., Jansee, A., Meinhof, C-G., Johnson, J., Giles, J. & Hamilton, S. (1995) Automating the mechanical subtraction of cDNA libraries. Presentation at the Third International Conference on Automation in Mapping and DNA Sequencing, Lawrence Berkeley National Laboratories, Berkeley California. November 5-8 1995.

Strauss, WM., Jaenisch, E & Jaenisch, R (1992) A strategy for rapid production and screening of yeast artificial chromosome libraries. Mamm. Genome 2: 150-157

Vidal, M., Boeke, J.D., Harlow, E. (1996a) Reverse two-hybrid system. WO 96/32503

Vidal, M., Brachmann, R.K., Fattaey, A., Harlow, E., Boeke, J.D. (1996b) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and protein-DNA interactions. Proc. Natl. Acad. Sci. USA, 93: 10315-10320

Wanker, E.E., Rovira, C., Scherzinger, E., Hasenbank, R., Waelter, S., Tait, D., Colicelli, J. and Lehrach H. (1997). HIP-I: a huntingtin interacting protein isolated by the yeast two-hybrid system. *Hum. Mol. Genet.* 6: 487-495.

Went, G., (1996) Quantitative Expression Analysis[SM] of cancer: new prospects for discovery and Therapy guidance[illegible]. Presentation at Advances in Gene Amplification & detection: New technology, Research & Clinical Applications. The Ritz-Carlton, McLean, Virginia. June 17-19

Wu, L.C., Wang, Z.W., Tsan, J.T., Spillman, M.A., Phung, A., Xu, X.L., Yang, M.C., Hwang, L.Y., Bowcock, A.M. and Baer, R. (1996). Identification of a RING protein that can interact in vivo with the BRCA1 gene product. *Nat. Genet.* **14**: 430-440.

Yang, M., Wu, Z. and Fields (1995). Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nucleic Acids Res.* **23**:1152-1156

Zhang, J. and Lautar, S. (1996). A yeast three-hybrid method to clone ternary protein complex components. *Anal. Biochem.* **242**:68-72.

Table 1

Oligonucleotide adapters for the construction of the novel yeast two-hybrid vectors pBTM118 a, b and c and pGAD428 a, b and c.

Oligonucleotide	Sequence (5'-3')
a sense	TCGAGTCGACGCGGCCGCTAA
A antisense	GGCCTTAGCGGCCGCGTCGAC
b sense	TCGAGGTCGACGCGGCCGCAGTAA
B antisense	GGCCTTACTGCGGCCGCGTCGACC
c sense	TCGAGAGTCGACGCGGCCGCTTAA
c antisense	GGCCTTAAGCGGCCGCGTCGACTC

Table 2

Two-hybrid vectors used for the expression of fusion proteins.

Plasmid	Fusion-protein	Insert (kb)	Counter-selection	Selection in yeast	Fusion protein Reference
PBTM117c	LexA	-	CAN1	TRP1	N/A
pBTM117c-HD1.6	LexA-HD1.6	1.6	CAN1	TRP1	Wanker et al., 1997
pBTM117c-HD3.6	LexA-HD3.6	3.6	CAN1	TRP1	Wanker et al., 1997
pBTM117c-SIM1	LexA-SIM1	1.1	CAN1	TRP1	Probst et al., 1997
pBTM117c-MJD	LexA-MJD	1.1	CAN1	TRP1	this work
pBTM117c-HIP1	LexA-HIP1	1.2	CAN1	TRP1	this work
PGAD427	GAL4ad	-	CYH2	LEU2	N/A
pGAD427-ARNT	GAL4ad-ARNT	1.4	CYH2	LEU2	Probst et al., 1997
pGAD427-HIP1	GAL4ad-HIP1	1.2	CYH2	LEU2	Wanker et al., 1997
pGAD427-HIPCT	GAL4ad-HIPCT	0.5	CYH2	LEU2	Wanker et al., 1997
pGAD427-14-3-3	GAL4ad-14-3-3	1.0	CYH2	LEU2	this work
pGAD427-LexA	Gal4ad-LexA	1.2	CYH2	LEU2	this work

Table 3

Yeast strains used for the 5-FOA counterselection and the automated interaction mating

Strain	Plasmids	Selected on
x1a	pBTM117c / pLUA	SD-trp-ade
x2a	pBTM117c-SIM1 / pLUA	SD-trp-ade
x3a	pBTM117c-HIP1 / pLUA	SD-trp-ade
y1α	pGAD427 / pLUA	SD-leu-ade
y2α	pGAD427-ARNT / pLUA	SD-leu-ade
y3α	pGAD427-LexA / pLUA	SD-leu-ade

Table 4

Identification of fusion proteins that activate the URA3 readout system.

a.

Strain	Plasmids	SD-trp -ade	SD-trp -ade+5-FOA	SD-trp -ade-ura
x1a	pBTM117c / pLUA	+	+	-
x2a	pBTM117c-SIM1 / pLUA	+	+	-
x3a	pBTM117c-HIP1 / pLUA	+	-	+

SD-trp-ade: Selective medium lacking tryptophan and adenine.
SD-trp-ade+5-FOA: Selective medium containing 0.2 % 5-FOA.
SD-trp-ade-ura: Selective medium lacking tryptophan, adenine and uracil.

b.

Strain	Plasmids	SD-leu -ade	SD-leu -ade+5-FOA	SD-leu -ade-ura
y1α	pGAD427 / pLUA	+	+	-
y2α	pGAD427 -ARNT/pLUA	+	+	-
y3α	pGAD427 -LexA/pLUA	+	-	+

SD-leu-ade: Selective medium lacking leucin and adenine.
SD-leu-ade+5-FOA: Selective medium containing 0.2 % 5-FOA.
SD-leu-ade-ura: Selective medium lacking leucin, adenine and uracil.

Table 5

Identification of fusion proteins that activate the LacZ readout system.

A. L40ccu yeast cells transformed with pBTM117c plasmid constructs expressing a fusion protein comprising the LexA DNA binding domain are plated on minimal medium lacking trptophan, buffered to pH 7.0 with potassium phosphate and containing 2 ug/ml of X-Gal (SD-trp/XGAL): Results for the state of the readout system for various auto-activating and non-auto-activating fusion proteins

Plasmid Construct	Fusion protein	Growth on SD-trp/XGAL	Blue colouration
pBTM117c	LexA	+	-
pBTM117c-HD1.6	LexA-HD1.6	+	-
pBTM117c-HD3.6	LexA-HD3.6	+	-
pBTM117c-SIM1	LexA-SIM1	+	-
pBTM117c-MJD	LexA-MJD	+	-
pBTM117c-HIP1	LexA-HIP1	+	+

B. L40ccuα yeast cells transformed with pGAD427 plasmid constructs expressing a fusion protein comprising the GAL4ad activation domain are plated on minimal medium lacking leucine, buffered to pH 7.0 with potassium phosphate and containing 2 ug/ml of X-Gal (SD-leu/XGAL): Results for the state of the readout system for various auto-activating and non-auto-activating fusion proteins.

Plasmid Construct	Fusion protein	Growth on SD-leu/XGAL	Blue colouration
pGAD427	GAL4ad	+	-
PGAD427-ARNT	GAL4ad-ARNT	+	-
PGAD427-HIP1	GAL4ad-HIP1	+	-
PGAD427-HIPCT	GAL4ad-HIPCT	+	-
PGAD427-14-3-3	GAL4ad-14-3-3	+	-
PGAD427-LexA	Gal4ad-LexA	+	+

CLAIMS

1. A method for the identification of at least one member of a pair or complex of interacting molecules, comprising:

(a) providing host cells containing at least two genetic elements with different selectable and counter-selectable markers, said genetic elements each comprising genetic information specifying one of said members, said host cells further carrying a readout system that is activated upon the interaction of said molecules;

(b) allowing at least one interaction, if any, to occur;

(c) selecting for said interaction by transferring progeny of said host cells to:

(ca) at least two different selective media, wherein each of said selective media allows growth of said host cells only in the absence of at least one of said counter-selectable markers and in the presence of a selectable marker; and

(cb) a further selective medium that allows identification of said host cells only on the activation of said readout system;

(d) identifying host cells containing interacting molecules that:

(da) do not activate said readout system on any of said selective media specified in (ca); and

(db) activate the readout system on said selective medium specified in (cb); and

(e) identifying at least one member of said pair or complex of interacting molecules

2. The method of claim 1 wherein said pair or complex of interacting molecules is selected from the group consisting of RNA-RNA, RNA-DNA, RNA-protein, DNA-DNA, DNA-protein, protein-protein, protein-peptide, or peptide-peptide interactions.

3. The method of claim 1 or 2 wherein said genetic elements are plasmids artificial chromosomes, viruses or other extra chromosomal elements.

4. The method of any one of claims 1 to 3 wherein said interactions lead to the formation of a functional transcriptional activator comprising a DNA-binding and a transactivating protein domain and which is capable of activating a responsive moiety driving the activation of said readout system.

5. The method of claim 4 wherein said readout system is a detectable protein.

6. The method of claim 5 wherein said detectable protein is encoded from at least one of the genes *lacZ*, HIS3, URA3, LYS2, *sacB*, *tetA*, gfp or HRPT.

7. The method of any one of claims 1 to 6 wherein said host cells are yeast cells, bacterial cells, mammalian cells, insect cells or plant cells.

8. The method of any one of claims 1 to 7 further comprising transforming or transfecting said host cells with said genetic elements prior to step (a).

9. The method of any one of claims 1 to 8 wherein cell fusion, conjugation or interaction mating is used for the generation of said host cells with said genetic elements prior to step (a).

10. The method of any one of claims 1 to 9 wherein said counter-selectable markers selected against in step (ca) are selected from the group of CAN1, CYH2, LYS2, URA3, *lacY*, *rpsL* HPRT and *sacB*.

11. The method of any one of claims 1 to 10 wherein said selectable marker is an auxotrophic or antibiotic marker.

12. The method of claim 11 wherein said auxotrophic or antibiotic marker is LEU2, TRP1, URA3, ADE2, HIS3, LYS2 or Zeocin.

13. The method of any one of claims 1 to 12 wherein progeny of host cells of step (b) are transferred to storage compartment.

14. The method of claim 13 wherein said transfer is effected or assisted by automation or a picking robot.

15. The method of claim 13 or 14 wherein said storage compartment comprises an anti-freeze agent.

16. The method of any one of claims 3 to 15 wherein said storage compartment is a microtiter plate.

17. The method of claim 16 wherein said microtiter plate comprises 384 wells.

18. The method of any one of claims 1 to 17 wherein said transfer in step (c) is made or assisted by automation, a spotting robot, pipetting or micropipetting device.

19. The method of claim 18 wherein said transfer is made to a planar carrier.

20. The method of claim 18 or 19 wherein said transfer is in a regular grid pattern of densities of 1 to 1000 clones per cm^2.

21. The method of any one of claims 18 to 20 wherein said planar carrier is a membrane.

22. The method of any one of claims 1 to 21 wherein said identification of said host cells in step (d) is effected by visual means from consideration of the activation state of said readout system.

23. The method of any one of claims 1 to 22 wherein said identification of said host cells in step (d) is effected by digital image storage, analysis or processing.

24. The method of any one of claims 1 to 23 wherein said identification of said at least one member of said pair of interacting molecules is effected by nucleic acid hybridization, antibody binding or nucleic acid sequencing.

25. The method of claim 24 wherein said identification made by said antibody reaction or said hybridization is effected using regular grids of said at least one member or of said genetic information encoding said at least one member.

26. The method of claim 25 wherein construction of said regular grids is effected by automation or a spotting robot.

27. The method of any one of claims 24 to 26 wherein said identification is effected by digital image storage, processing or analysis.

28. The method of any one of claims 24 to 27 wherein nucleic acid molecules, prior to said identification, are amplified by PCR or are amplified in as a part of said genetic elements, preferably in bacteria and most preferably in E.coli.

29. The method of any one of claims 1 to 28 wherein, prior to step (a) a preselection against clones that express a single molecule able to activate the readout system is carried out on culture media comprising a counterselective compound.

30. The method of claim 29 wherein said counterselective compound is 5-fluoro orotic acid, canavanine, cycloheximide, sucrose,tONPG, streptomycin or α-amino-adipate.

31. A method for the production of a pharmaceutical composition comprising formulating said at least one member of the interacting molecules identified by the method of any one of claims 1 to 30 in a pharmaceutically acceptable form.

32. A method for the production of a pharmaceutical composition comprising formulating an inhibitor of the interaction of the interacting molecules identified by the method of any one of claims 1 to 30 in a pharmaceutically acceptable form.

33. A method for the production of a pharmaceutical composition comprising identifying a further molecule of a cascade of interacting molecules, of which the at least one member of said interacting molecules identified by the method of any one of claims 1 to 30 is a part of or identifying an inhibitor of said further molecule.

34. Kit comprising at least one of the following:

(f) host cells as identified in any of the preceding claims and at least one genetic element comprising said genetic information specifying at least one of said possibly interacting molecules containing a counterselectable marker and specified in any of the preceding claims;

(g) host cells as identified in any of the preceding claims and at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified in any of the preceding claims;

(h) at least one genetic element comprising said genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified in any of the preceding claims;

(i) at least one genetic element not comprising genetic information specifying at least one of said potentially interacting molecules containing a counterselectable marker and specified in any of the preceding claims;

(j) host cells comprising at least one and preferably at least two of said genetic elements specified in (h) or (i);

(k) at least one planar carrier carrying nucleic acid or protein from said host cells comprising at least one member of said genetic elements specified in any of the preceding claims wherein said nucleic acid or protein is affixed to said carrier in grid form and optionally solutions to effect hybridization or binding of nucleic acid probes or proteins to said molecules affixed to said grid;

(l) at least one storage compartment, planar carrier or computer disc comprising or/and characterizing

genetic elements, host cells, storage compartments or carriers identified in any of (f) to (k); and/or

(m) at least one yeast strain comprising a *can1* and a *cyh2* mutation.

35. The kit of claim 34, wherein said host cells of (f), (g) or (j) are contained in at least one storage compartment.

36. The kit of claim 34 or 35, wherein said genetic information or said potentially interacting molecules encoded by said genetic information as specified in (i) or (iii) are contained in at least one storage compartment.

37. A computer implemented method for, storing and analysing data relating to potential members of at least one pair or complex of interacting molecules encoded by nucleic acids originating from biological samples, said methods comprising

(n) retrieving from a first data-table information for a first nucleic acid, wherein said information comprises:

(oa) a first combination of letters and/or numbers uniquely identifying the nucleic acid, and

(ob) the type of genetic element comprising said nucleic acid and

(oc) a second combination of letters and/or numbers uniquely identifying a clone in which a potential member encoded by said nucleic acid was tested for interaction with at least one other potential member of a pair or complex of interacting molecules

(p) using said second combination of letters and/or numbers to retrieve from said first data-table or optionally further data-tables, information identifying additional nucleic acids encoding for said at least one other potential member in step a3).

38. The method of claim 37 further comprising, using said second combination of letters and/or numbers in step (oc) to retrieve from a second data-table further information, where said further information at least comprises the interaction class of said clone, and optionally additional information comprising,
(q) the physical location of the clone,
(r) predetermined experimental details pertaining to creation of said clone, including at least one of,
(ra) tissue, disease-state or cell source of the nucleic acid,
(rb) cloning details, and
(rc) membership of a library of other clones,

39. The method of claim 38 further comprising, using said information of step (o) on said first and/or of step (p) on additional nucleic acids to relate to a third data-table further characterising said first and/or additional nucleic acids, where said further characterising comprises at least one of

(s) hybridization data;
(t) oligonucleotide fingerprint data;
(u) nucleotide sequence;
(v) in-frame translation of the said nucleic acids;
(w) tissue, disease-state or cell source gene expression data; and

optionally identifying the protein domain encoded by said first or additional nucleic acids.

40. The method of claim 39 further comprising, identifying if said potential members encoded by the nucleic acids interact, by considering said interaction class of said clone in which nucleic acids were tested for said interaction in step (oc)

41. The method of one of claims 37 to 40, wherein said data relates to 10 to 100 potential members, preferably 100 to 1000 potential members., more preferably 1000 to 10000 potential members and most preferably more than 10,000 potential members.

42. The method of one of claims 37 to 41, wherein said data was generated by the method of claims 1 to 36.

43. The method of claims 38 to 42, wherein said interaction class comprises one of the following:
(x) Positive
(y) Negative
(z) False Positive

44. The method of one of claims 40 to 43 wherein sticky proteins are identified by consideration of the number of occurrences a given member is identified to interact with many different members in different clones of said positive interaction class.

45. The method of one of claims 37 to 44, wherein said first data-table forms part of a first database, and said second and third data tables form part of at least a second database.

46. The method of claim 45, wherein said second database is held on a computer readable memory separate from the computer readable memory holding said first database, and said database is accessed via a data exchange network.

47. The method of claim 46, wherein said second database comprises nucleic acid or protein sequence, secondary or tertiary structure, biochemical, biographical or gene expression information.

48. The method of claims 37 to 47, wherein data entry to said first, second or further data tables is controlled automatically from said first data base by access to other computer data, programs or computer controlled robots.

49. The method of one of claims 37 to 48, wherein at least one workflow management system is built around particular sets of data to assist in the progress of the method of claims 1 to 36.

50. The method of claim 49, wherein said workflow management system is software to assist in the progress of the identification of members of a pair or complex of interacting molecules using the method of hybridization as specified in claims 24 to 28

51. The method of claims 37 to 50, wherein said data are investigated by queries of interest to an investigator.

52. The method of claim 51, wherein said queries include at least one of,

(aa) identifying the interaction or interaction pathway between a first and second member of an interaction network

(ab) identifying the interaction pathway between a first and second member of an interaction network and through at least one further member,

(ac) identifying the interaction or interaction pathway between at least two members characterised by nucleotide acid or protein sequences, secondary or tertiary structures, and

(ad) identifying interactions or interaction pathways that are different for said different tissue, disease-state or cell source.

53. The method of claims 51 or 52, wherein parts of said information is stored in a controlled format to assist

data query procedures.

54. The method of claims 51 to 53, wherein the results of said queries are displayed to the investigator in a graphical manner.

55. The method of claims 54, wherein a sub-set of data comprising data characterising nucleic acids identified as encoding members of a pair or complex of interacting molecules of claim 40 is stored in a further data-table or data base.

56. The method of claim 55 wherein consideration of the number of occurrences a given member is identified to interact with a second or further member is used to decide if said data characterising nucleic acids form part of said sub-set of data.

57. The method of claims 55 or 56, wherein additional information or experimental data is used to select those data to form part of said subset.

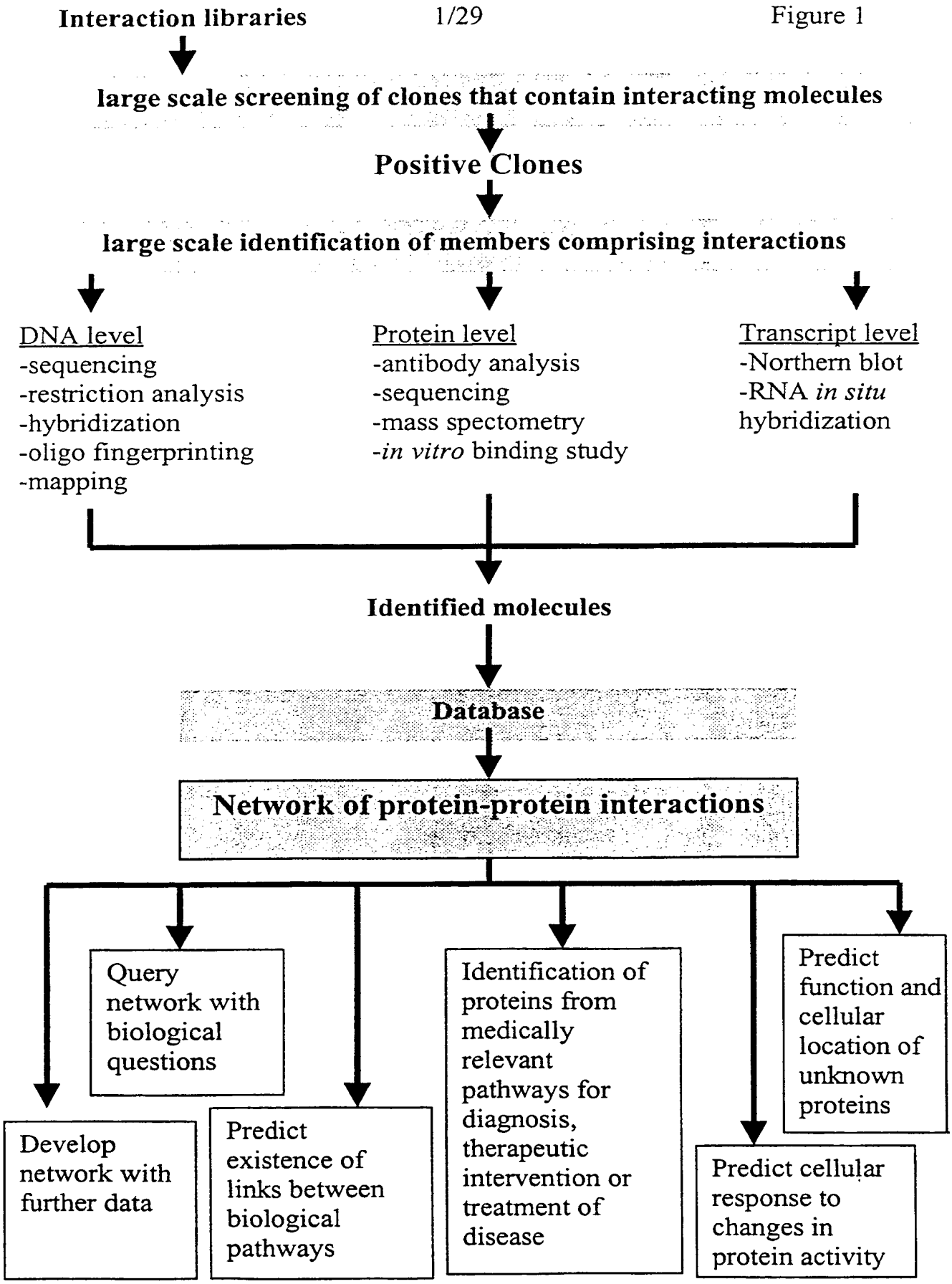
58. The method of claims 55 to 57, wherein to speed certain data query procedures, the structure in which the data is stored in the computer readable memory is modified.

59. The method of one of claims 37 to 58, wherein the data is held in relational or object oriented data bases.

60. A data storage scheme comprising a data table that holds information on each member of an interaction, where a record in said table represents each member of an interaction, and in which members are indicated to form interactions by sharing a common name.

61. The data storage scheme of claim 60, wherein said common name is a clone name or unique combination of letters and/or numbers comprising said clone name.

Figure 1



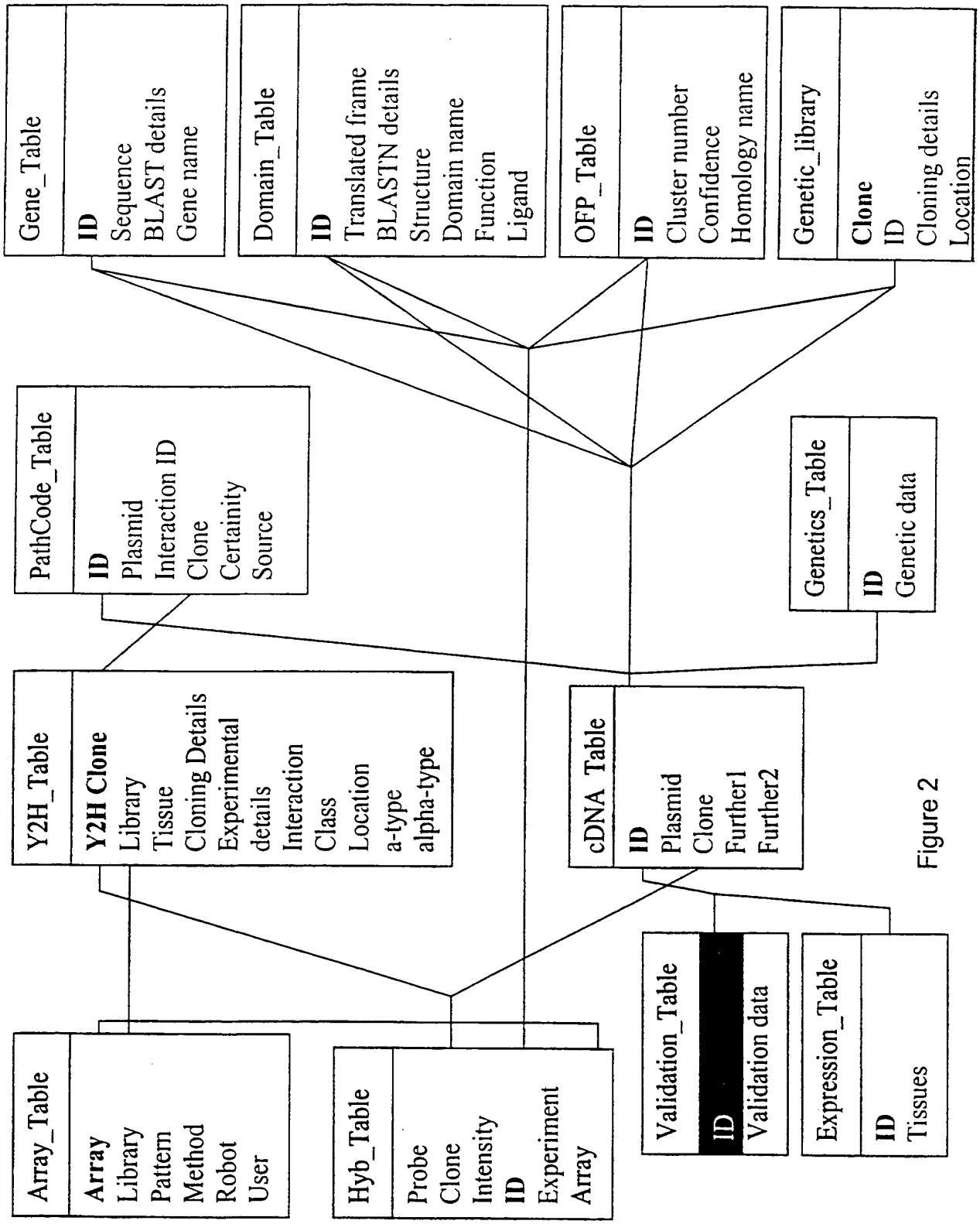
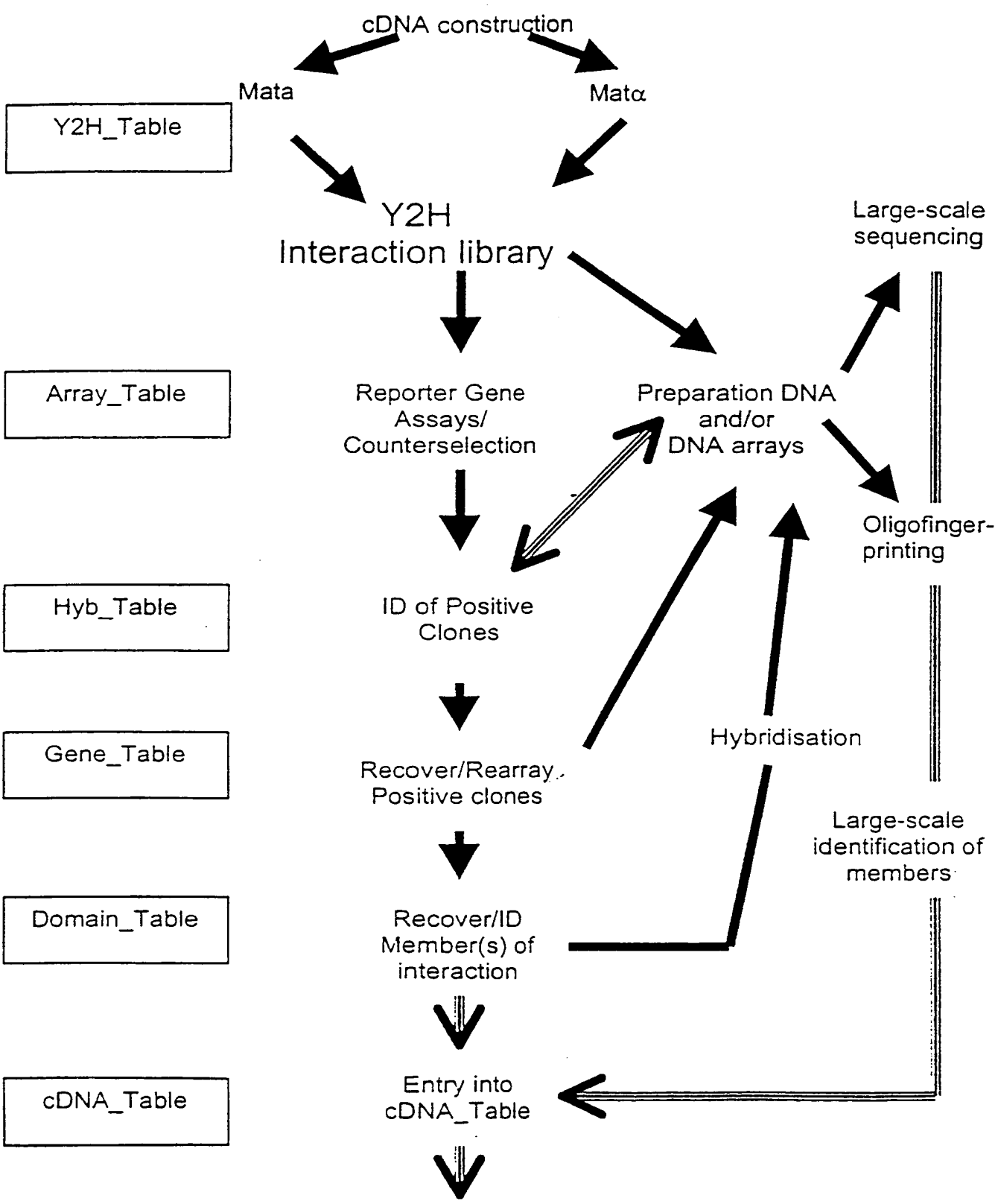

Gene_Table
ID
Sequence
BLAST details
Gene name
Domain_Table
ID
Translated frame
BLASTN details
Structure
Domain name
Function
Ligand
OFP_Table
ID
Cluster number
Confidence
Homology name
Genetic_library
Clone
ID
Cloning details
Location
PathCode_Table
ID
Plasmid
Interaction ID
Clone
Certainity
Source
Genetics_Table
ID
Genetic data
Y2H_Table
Y2H Clone
Library
Tissue
Cloning Details
Experimental details
Interaction
Class
Location
a-type
alpha-type
cDNA_Table
ID
Plasmid
Clone
Further1
Further2
Array_Table
Array
Library
Pattern
Method
Robot
User
Hyb_Table
Probe
Clone
Intensity
ID
Experiment
Array
Validation_Table
ID
Validation data
Expression_Table
ID
Tissues

Figure 2



(continuation from the previous page)

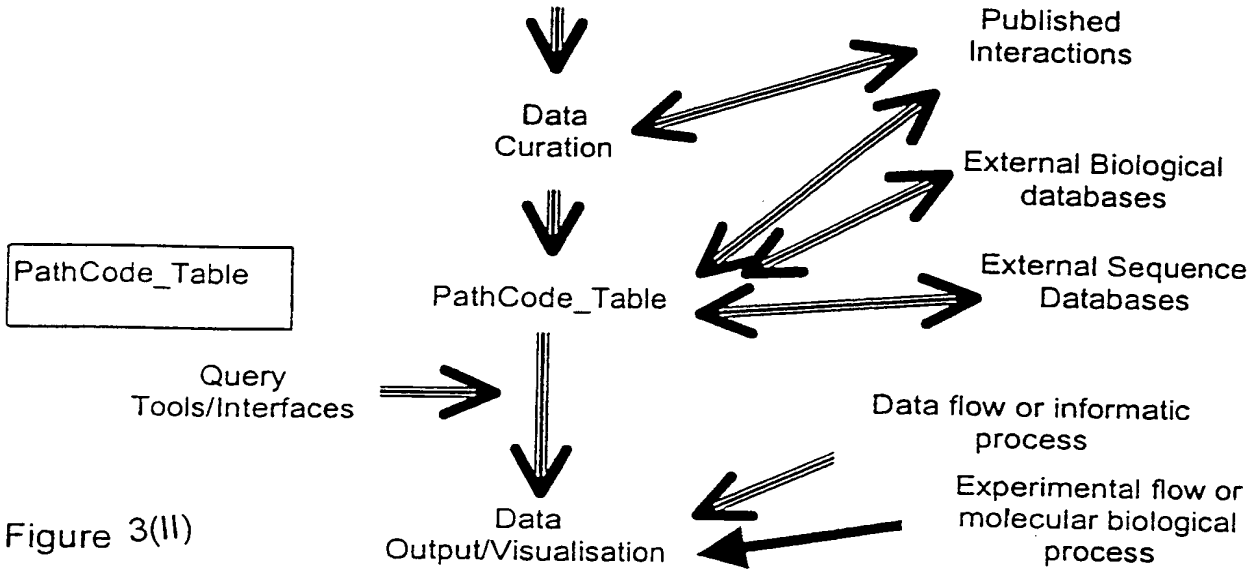
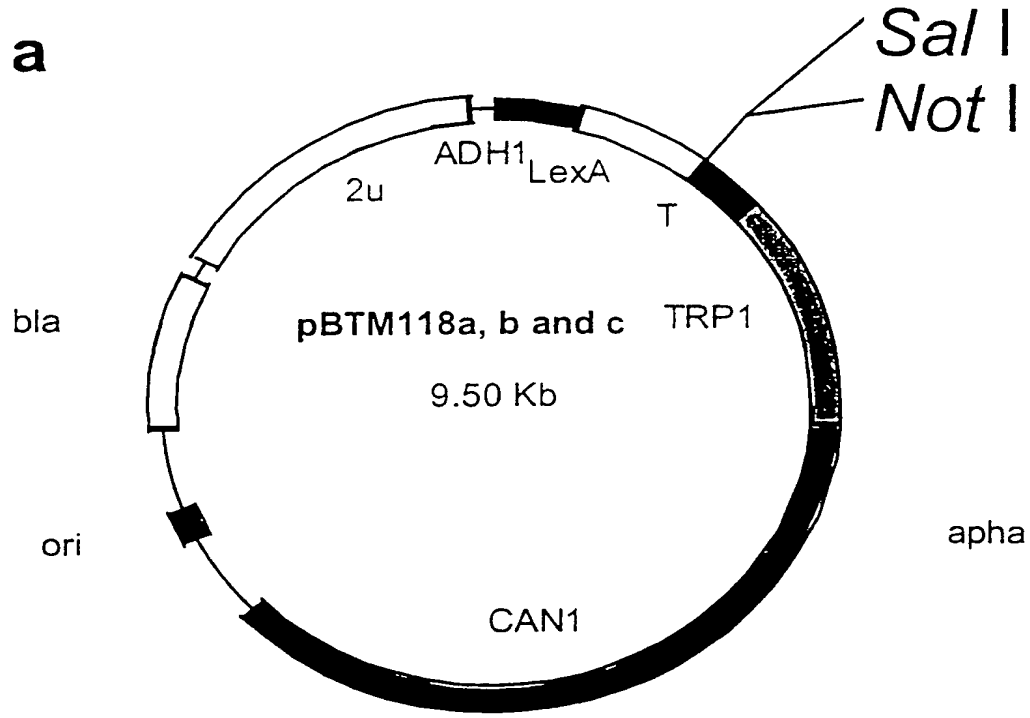


Figure 3(II)


a
Sal I
Not I
ADH1
LexA
2u
T
bla
pBTM118a, b and c
TRP1
9.50 Kb
ori
apha
CAN1

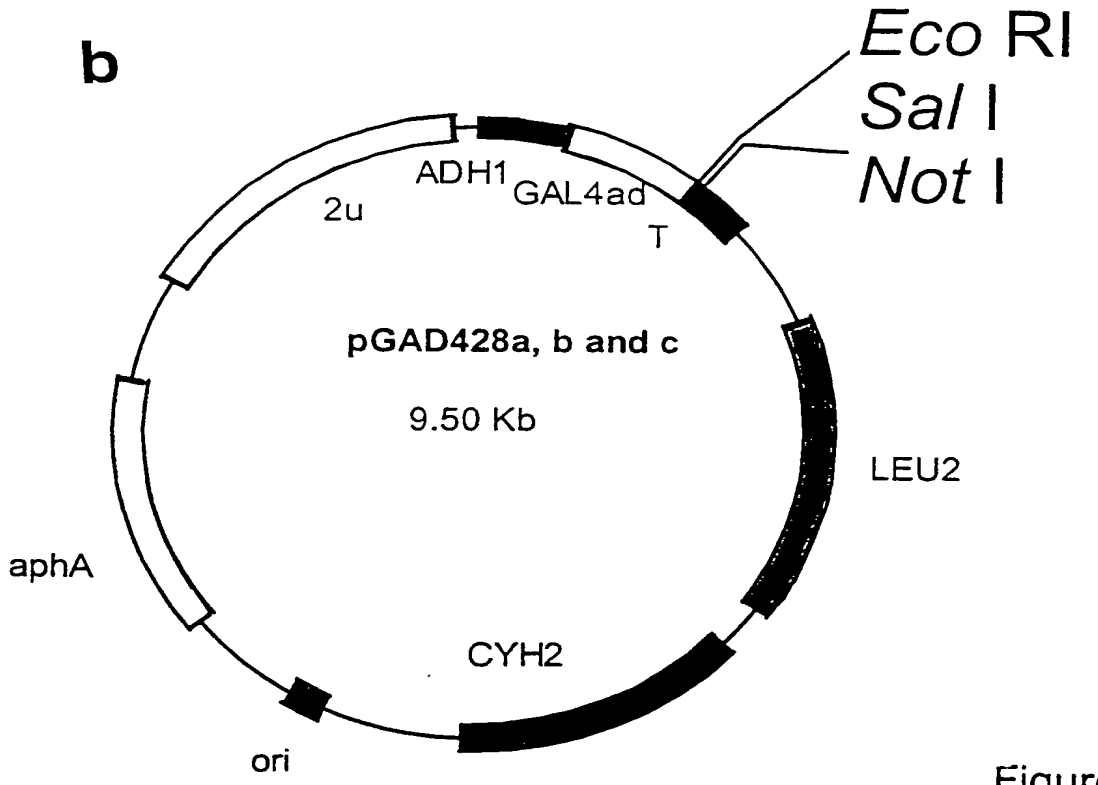

b
Eco RI
Sal I
Not I
ADH1
GAL4ad
2u
T
pGAD428a, b and c
9.50 Kb
LEU2
aphA
CYH2
ori

Figure 4 (I)

b

a) TCG AGT CGA CGC GGC CGC TAA CCG G

Sal I *Not* I STOP

b) TCG AGG TCG ACG CGG CCG CAG TAA CCG G

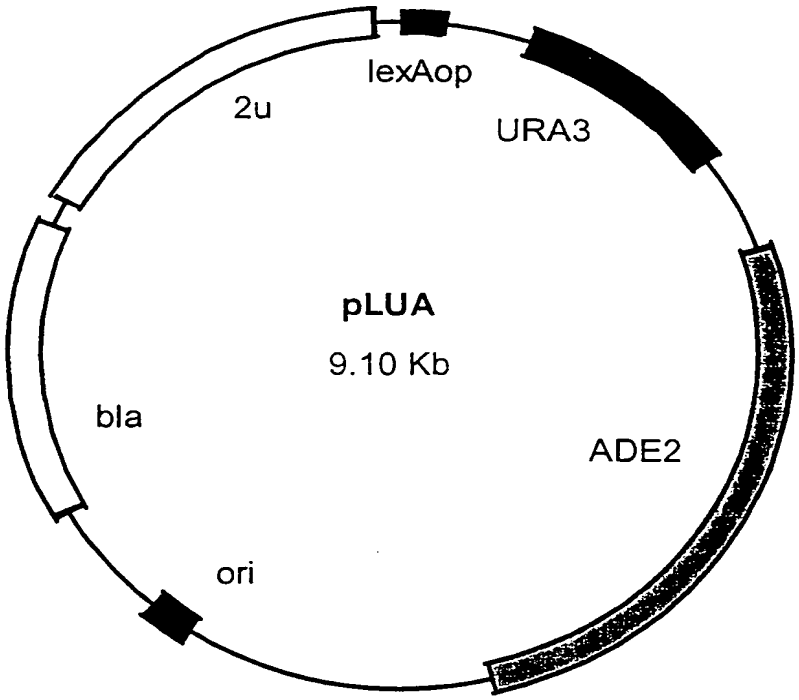
Sal I *Not* I STOP

c) TCG AGA GTC GAC GCG GCC GCT TAA CCG G

Sal I *Not* I STOP

Figure 4 (II)

Figure 5



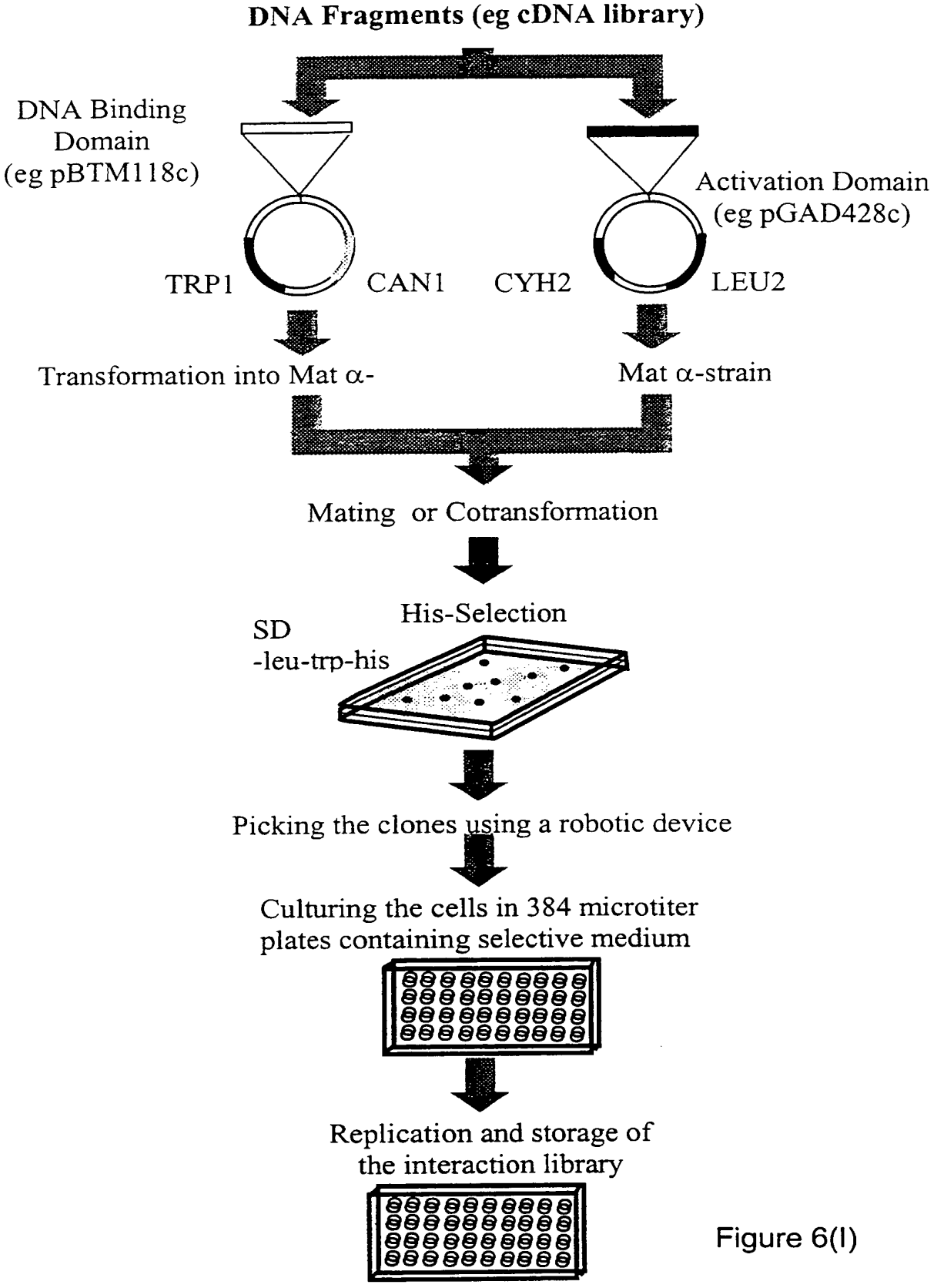

DNA Fragments (eg cDNA library)
DNA Binding Domain (eg pBTM118c)
Activation Domain (eg pGAD428c)
TRP1
CAN1
CYH2
LEU2
Transformation into Mat α-
Mat α-strain
Mating or Cotransformation
His-Selection
SD -leu-trp-his
Picking the clones using a robotic device
Culturing the cells in 384 microtiter plates containing selective medium
Replication and storage of the interaction library

Figure 6(I)

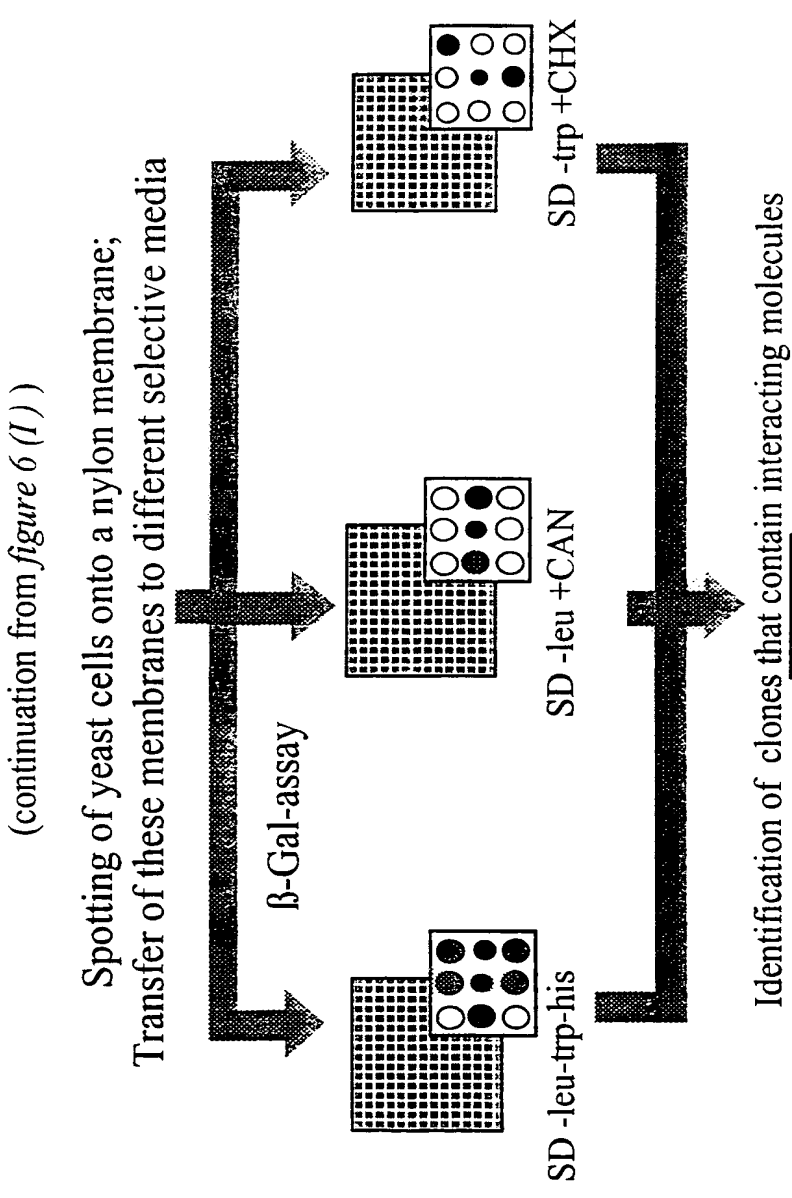

(continuation from figure 6 (I))
Spotting of yeast cells onto a nylon membrane;
Transfer of these membranes to different selective media
ß-Gal-assay
SD -leu-trp-his
SD -leu +CAN
SD -trp +CHX
Identification of clones that contain interacting molecules
(continuation see figure 6(III))

Figure 6 (II)

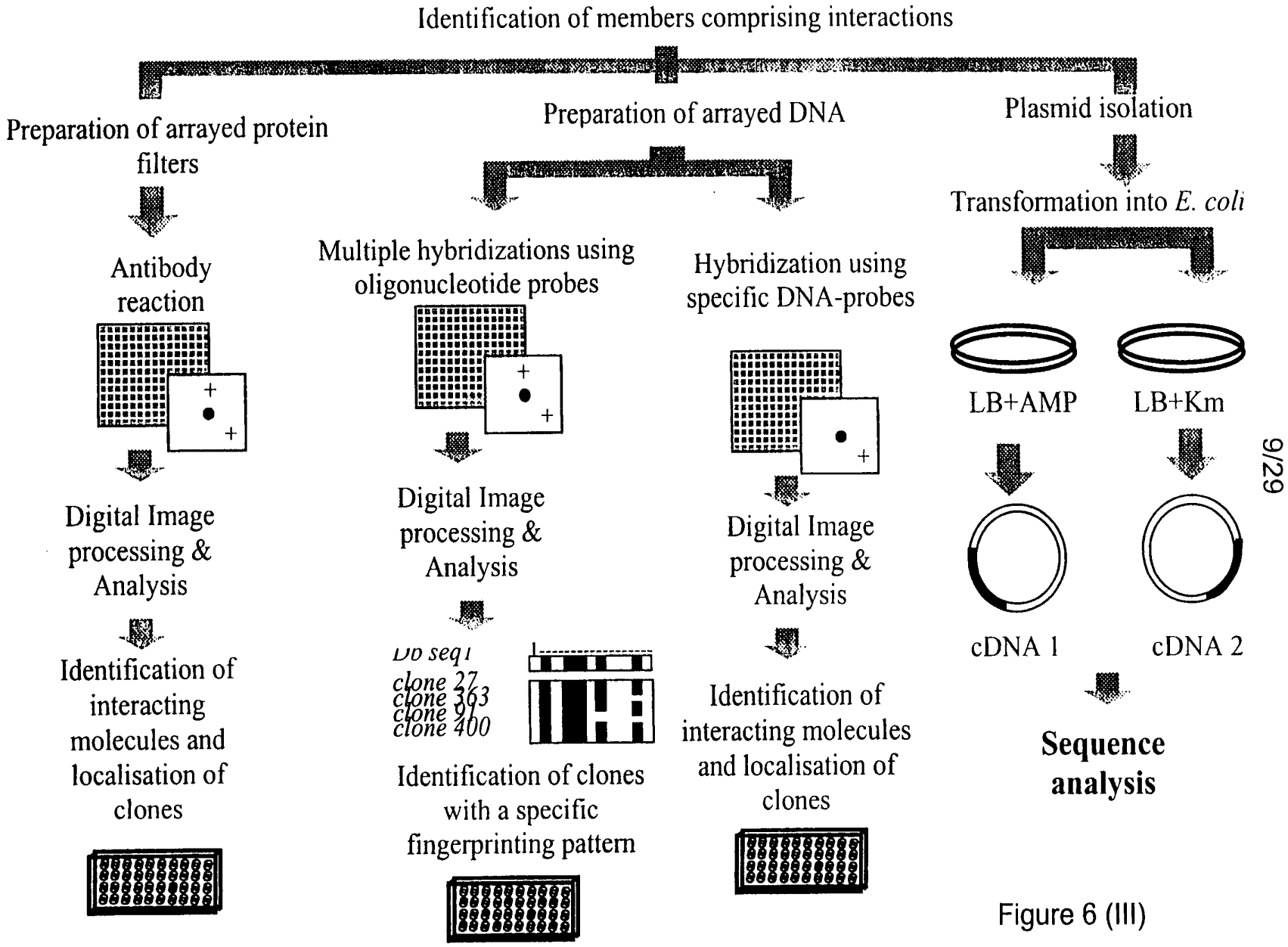


Figure 6 (III)

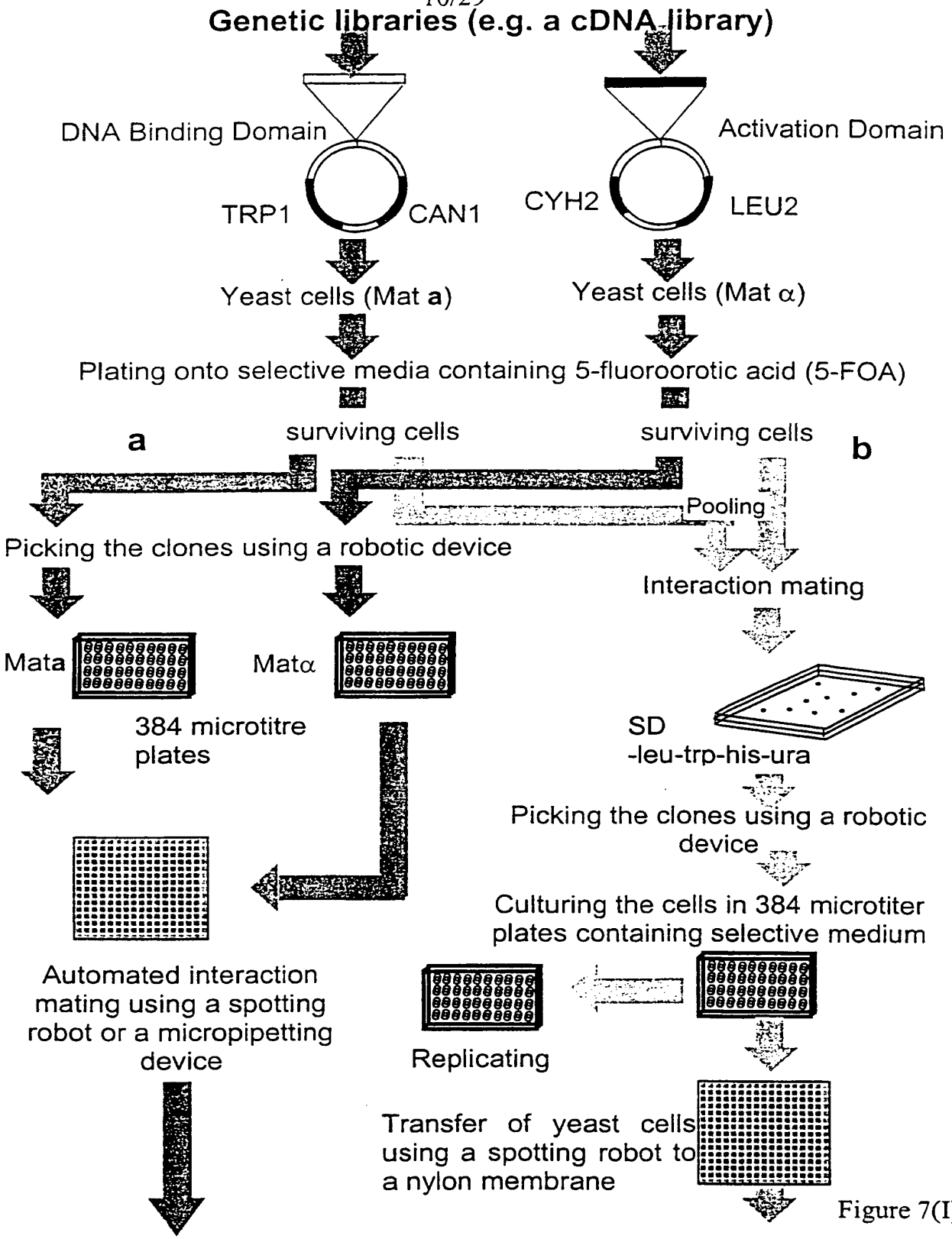

Genetic libraries (e.g. a cDNA library)
DNA Binding Domain
TRP1
CAN1
Activation Domain
CYH2
LEU2
Yeast cells (Mat a)
Yeast cells (Mat α)
Plating onto selective media containing 5-fluoroorotic acid (5-FOA)
surviving cells
surviving cells
a
b
Pooling
Picking the clones using a robotic device
Interaction mating
Mata
Matα
384 microtitre plates
SD
-leu-trp-his-ura
Picking the clones using a robotic device
Culturing the cells in 384 microtiter plates containing selective medium
Automated interaction mating using a spotting robot or a micropipetting device
Replicating
Transfer of yeast cells using a spotting robot to a nylon membrane

Figure 7(I)

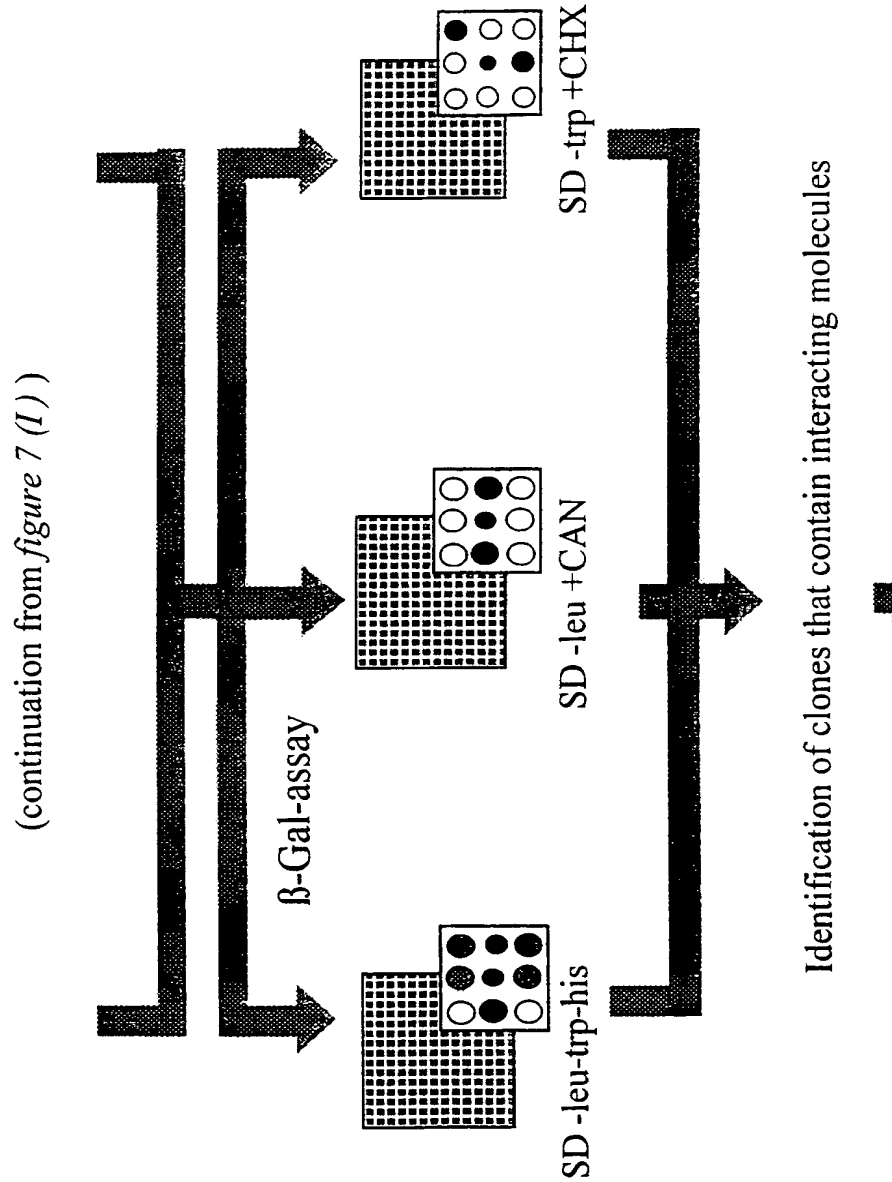


Figure 7 (II)

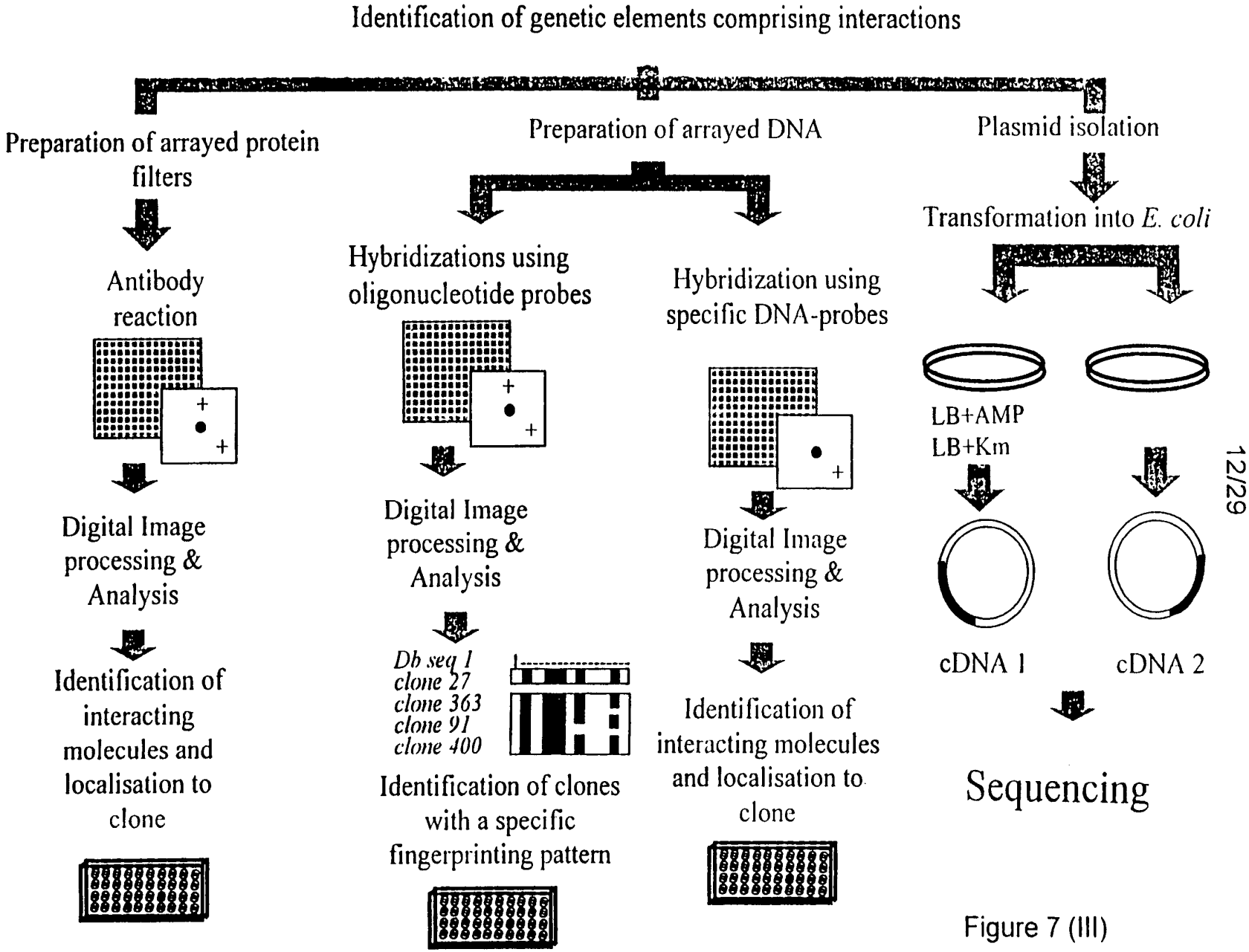


Figure 7 (III)

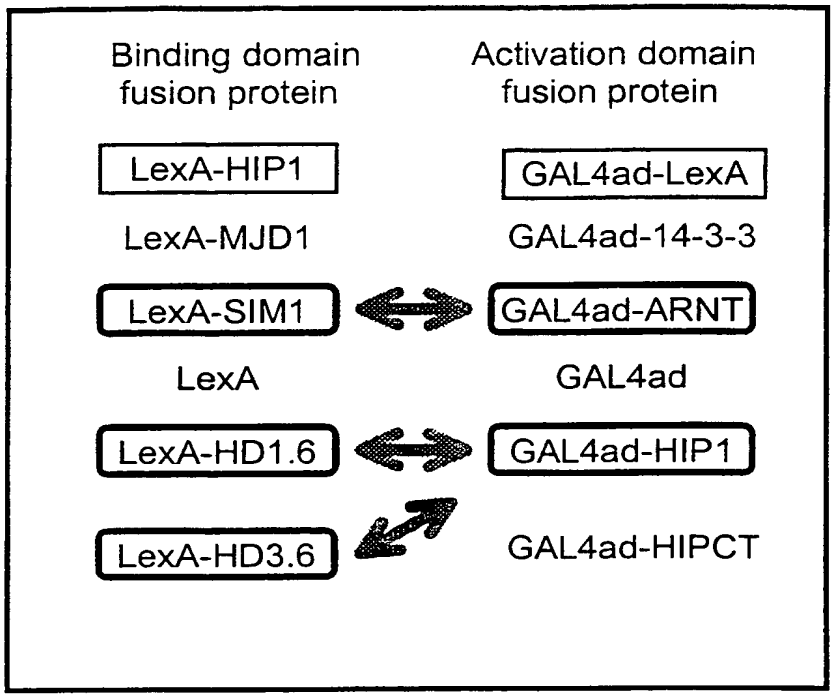


Figure 8

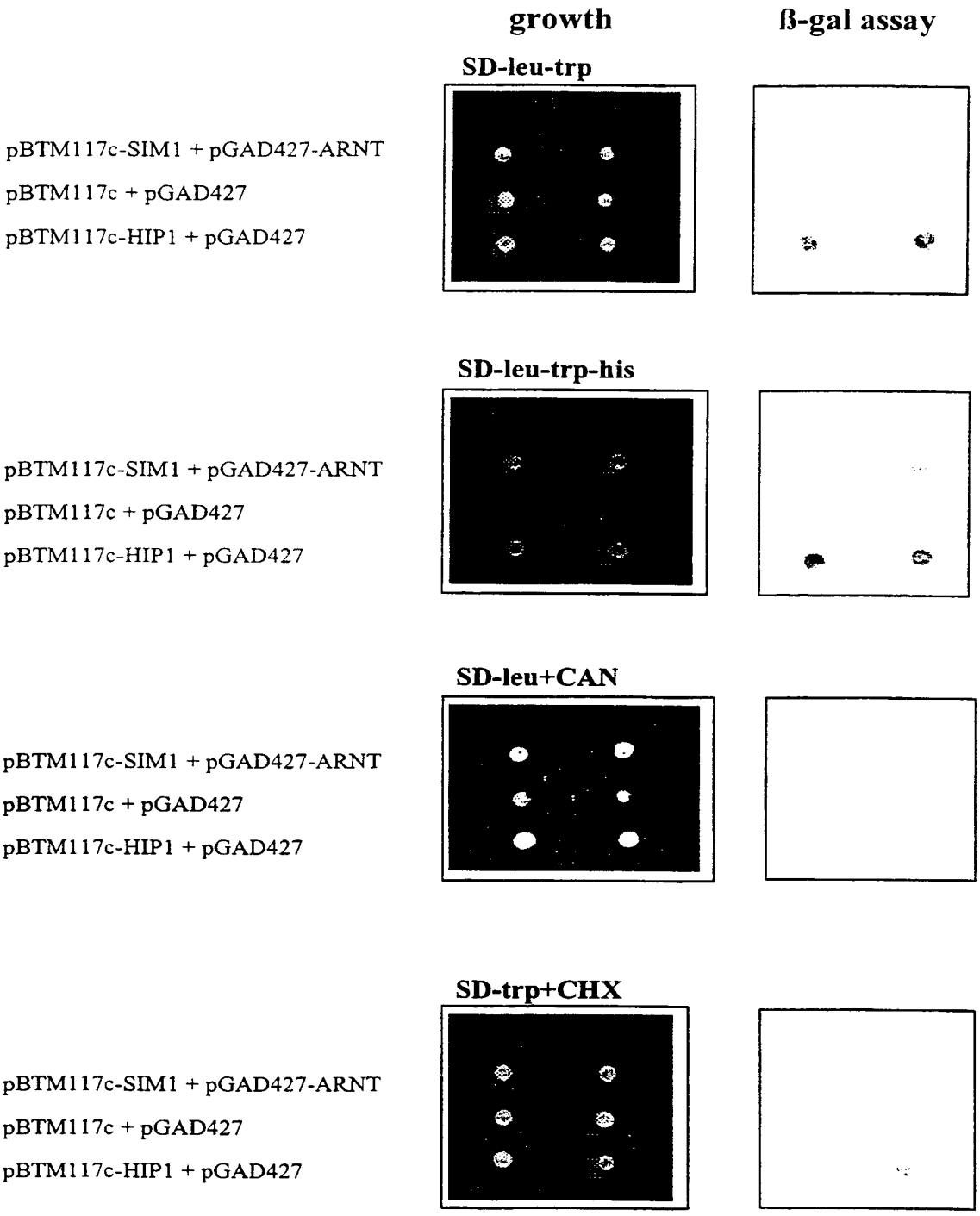
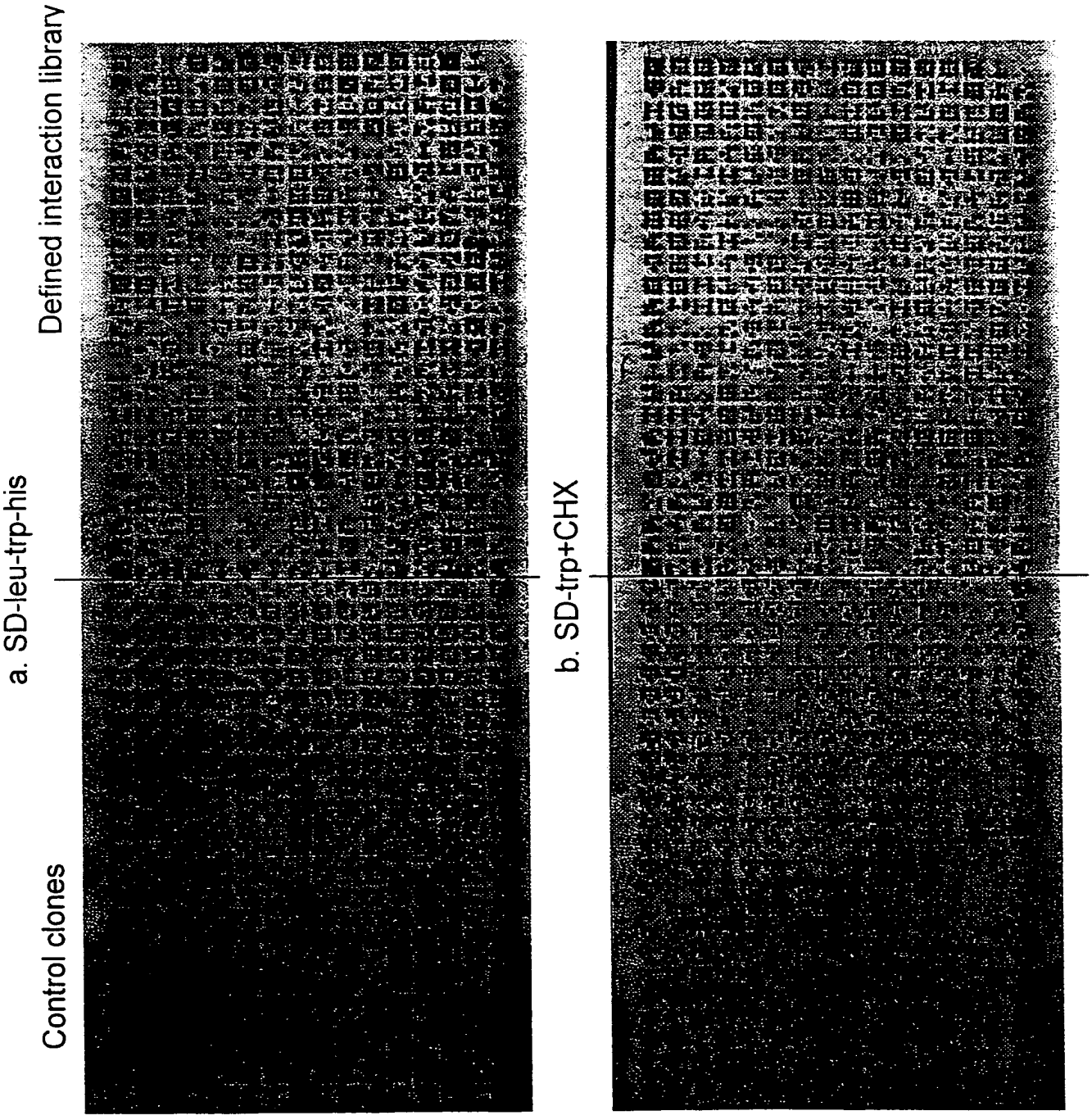

growth
ß-gal assay
SD-leu-trp
pBTM117c-SIM1 + pGAD427-ARNT
pBTM117c + pGAD427
pBTM117c-HIP1 + pGAD427
SD-leu-trp-his
pBTM117c-SIM1 + pGAD427-ARNT
pBTM117c + pGAD427
pBTM117c-HIP1 + pGAD427
SD-leu+CAN
pBTM117c-SIM1 + pGAD427-ARNT
pBTM117c + pGAD427
pBTM117c-HIP1 + pGAD427
SD-trp+CHX
pBTM117c-SIM1 + pGAD427-ARNT
pBTM117c + pGAD427
pBTM117c-HIP1 + pGAD427

Figure 9

Figure 10 (I)



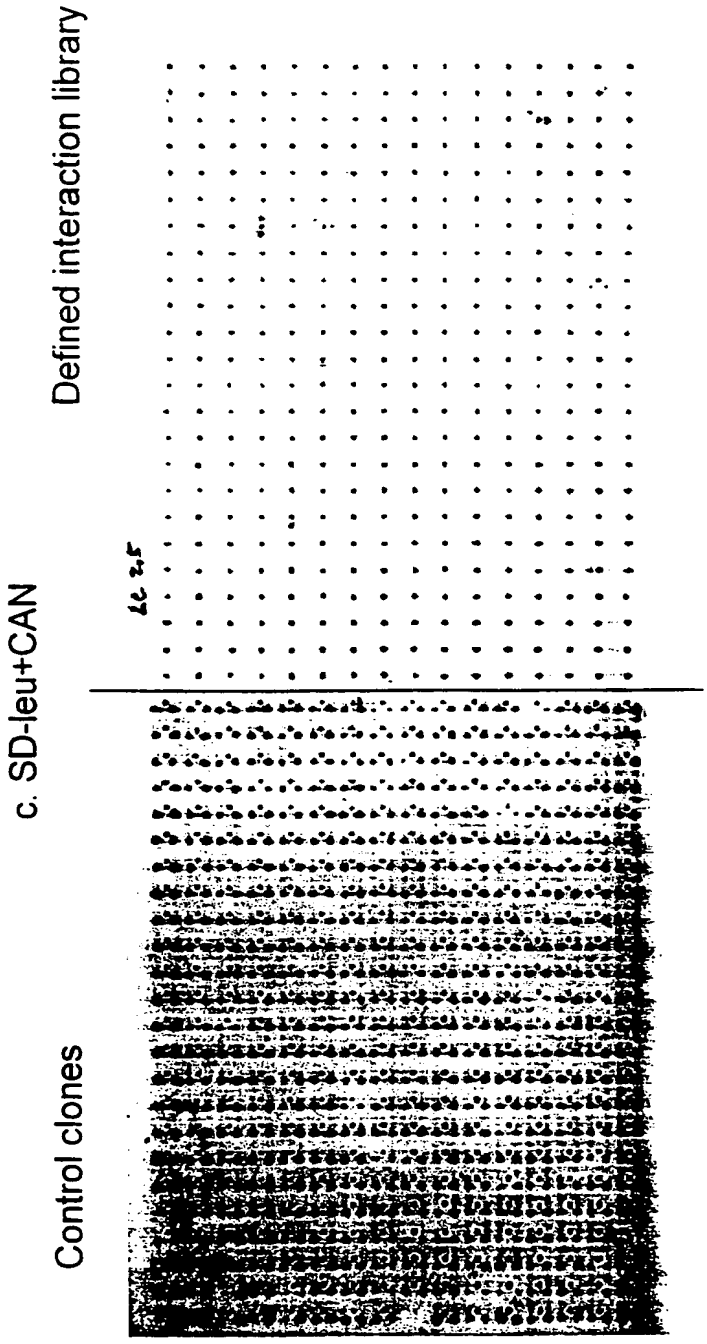


Figure 10 (II)

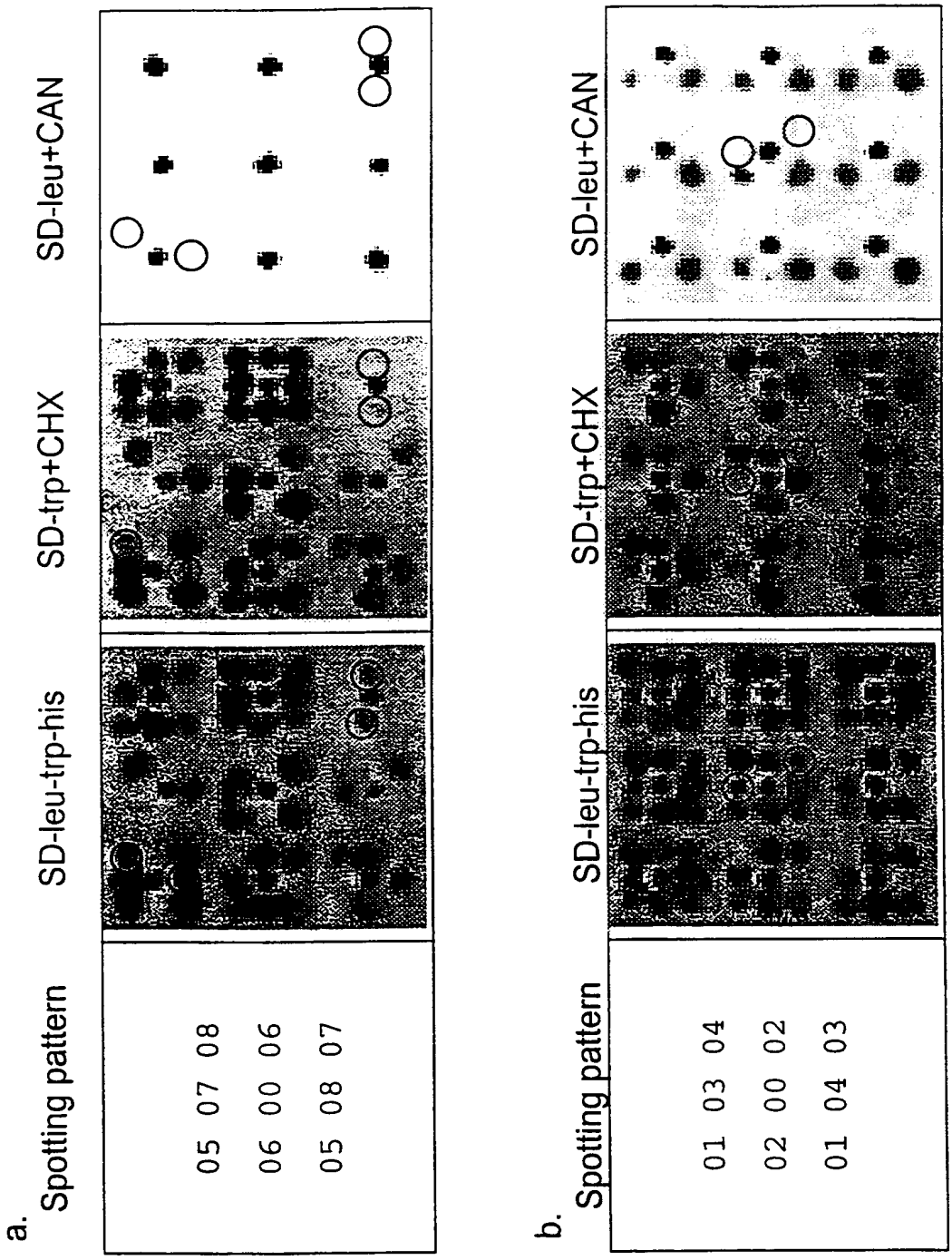


Figure 11

a. 18/29

GPC - PositiveFind

	Clone	SD-leu-trp-his	SD-trp+CHX	SD-leu+CAN
675	2O8	2	0	0
687	2O22	2	0	0
689	2P1	3	0	0
690	2P2	2	0	0
693	2P5	3	0	0
696	2P8	3	0	0
699	2P12	3	0	0
709	2P23	2	0	0
710	2P24	2	0	0
711	3A2	3	0	0
712	3A3	3	0	0
713	3A4	3	0	0
714	3A5	3	0	0
716	3A7	3	0	0
717	3A8	3	0	0
718	3A11	3	0	0
719	3A12	2	0	0
720	3A14	3	0	0
721	3A15	3	0	0
722	3A18	2	0	0
723	3A21	2	0	0
724	3A23	2	0	0
725	3A24	3	0	0
727	3B2	3	0	0
728	3B3	3	0	0
729	3B4	3	0	0
730	3B5	3	0	0
731	3B6	3	0	0
732	3B7	3	0	0
733	3B8	3	0	0
734	3B9	3	0	0
735	3B10	3	0	0
736	3B12	3	0	0
737	3B13	3	0	0
738	3B14	3	0	0
739	3B15	3	0	0
741	3B17	2	0	0

b.

GPC - PositiveFind

1	Clone	SD-leu-trp-his	SD-trp+CHX	SD-leu+CAN
1	Clone	SD-leu-trp-his	SD-trp+CHX	SD-leu+CAN
1733	6L22	3	0	0
2302	8N24	2	0	0

Figure 12

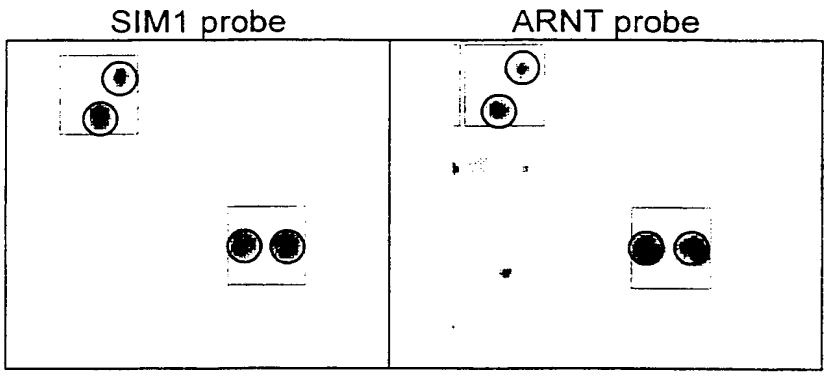


Figure 13

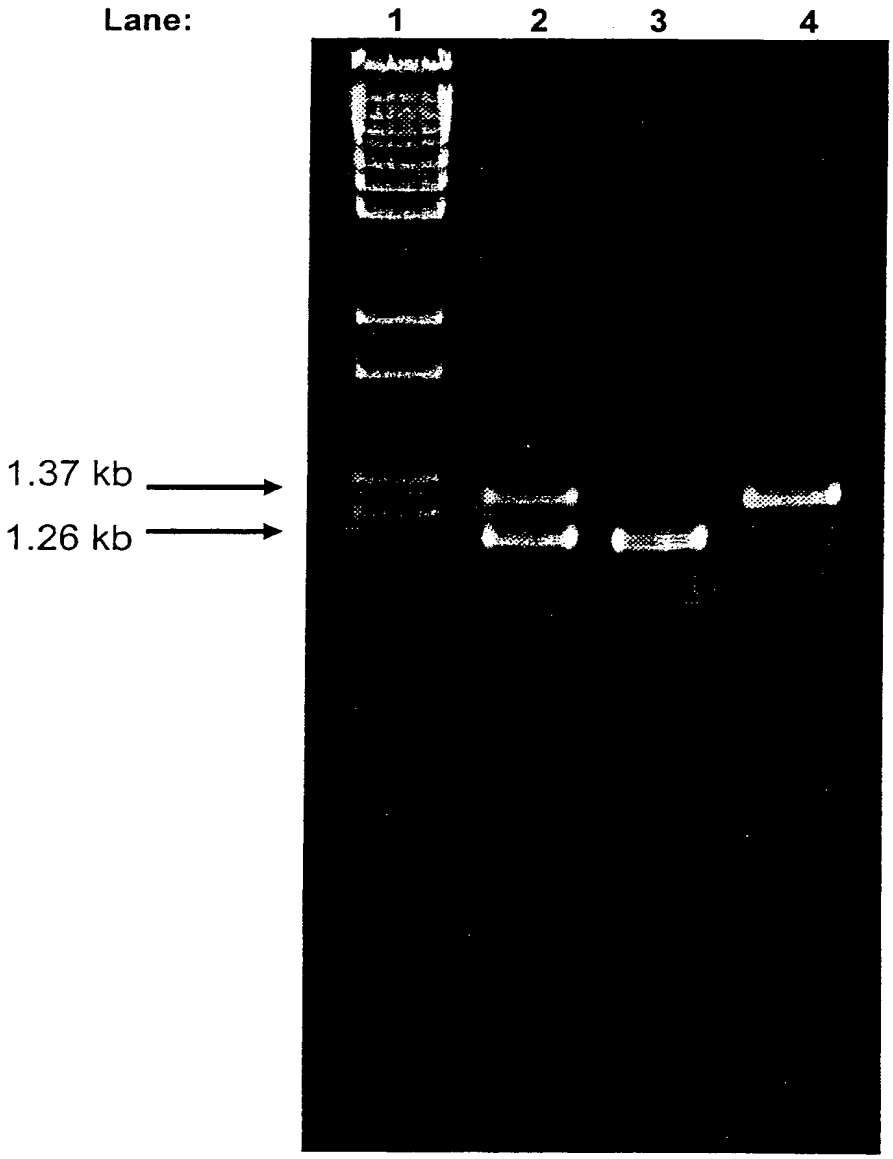


Figure 14

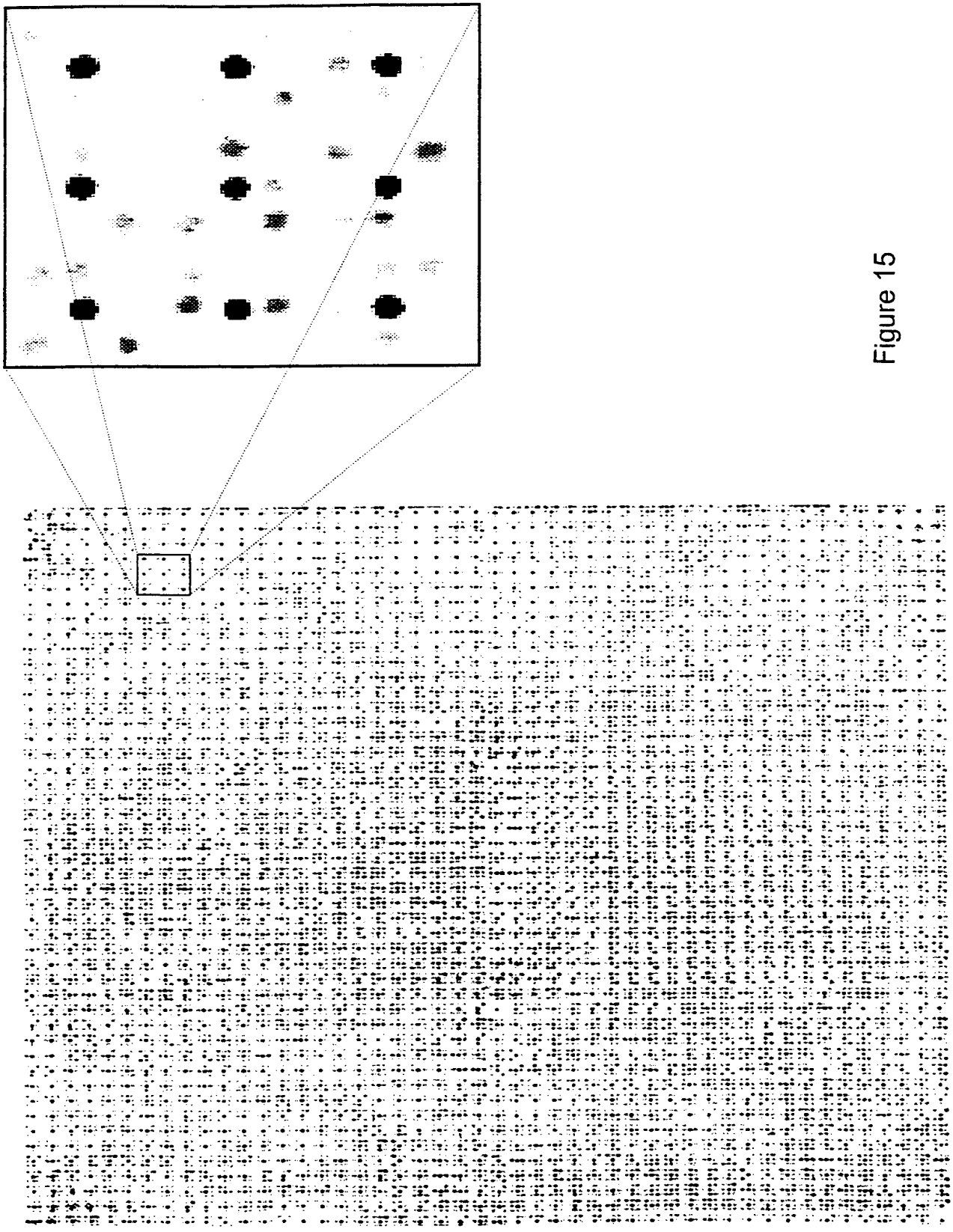
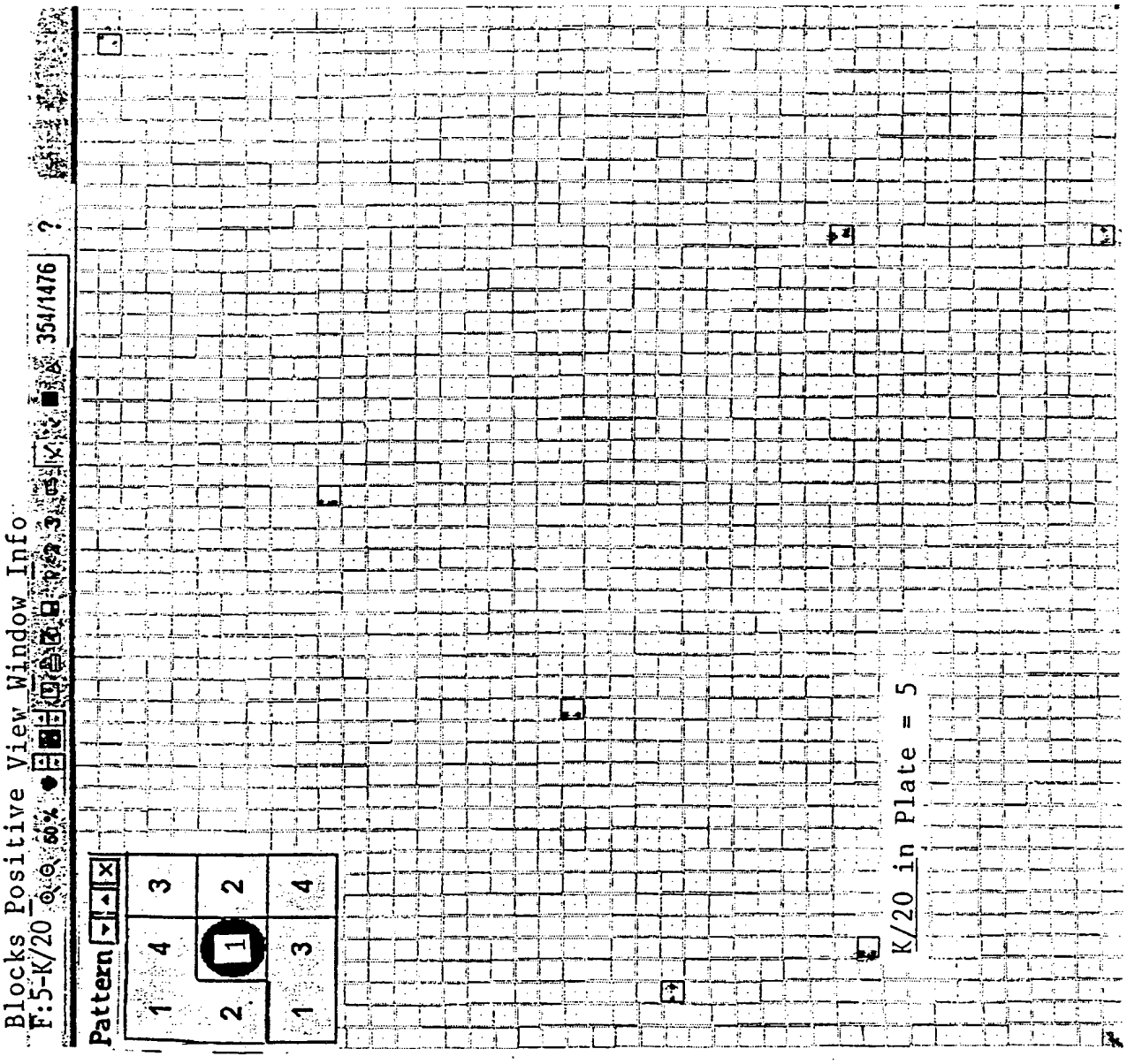


Figure 15

Figure 16



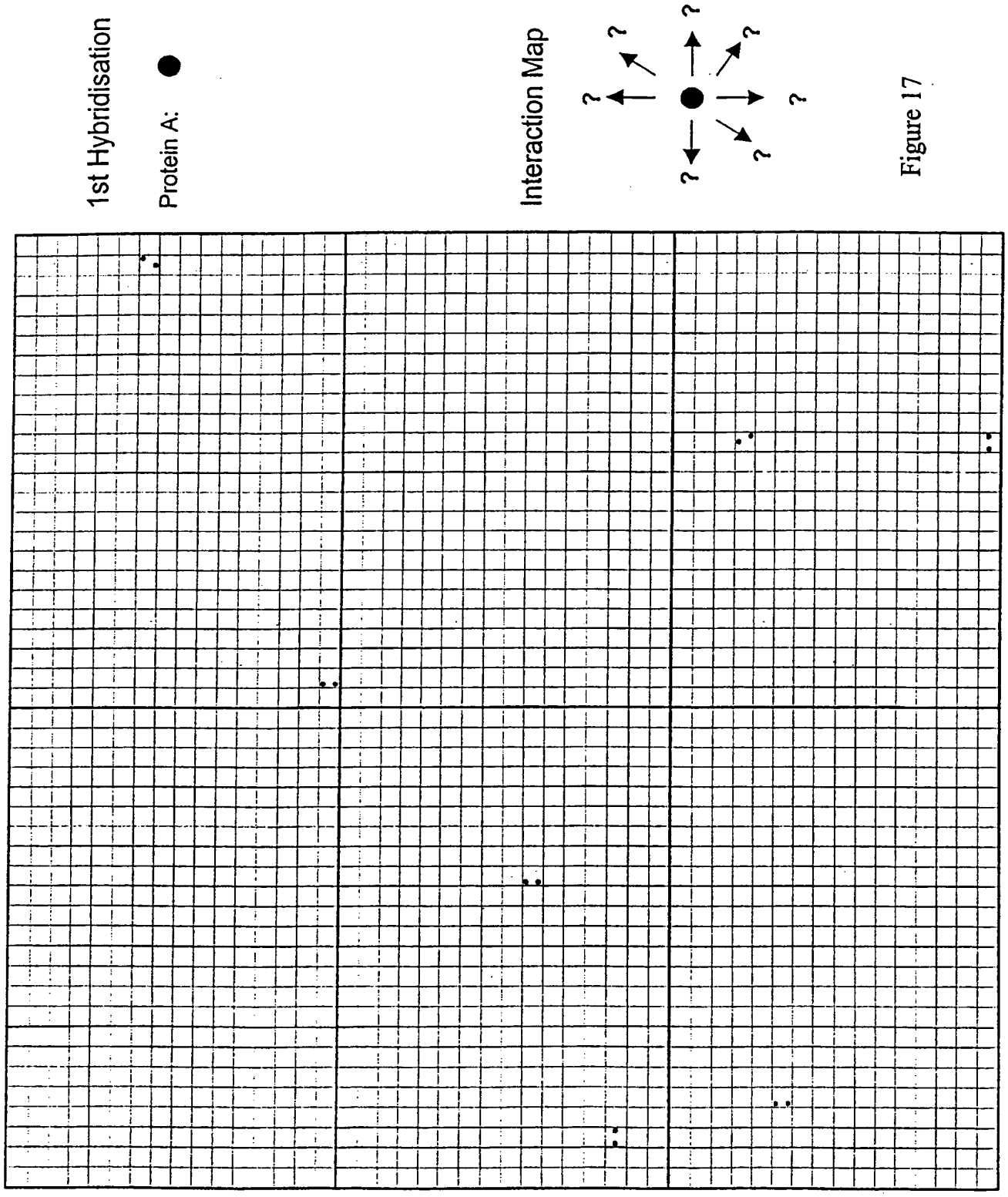


Figure 17

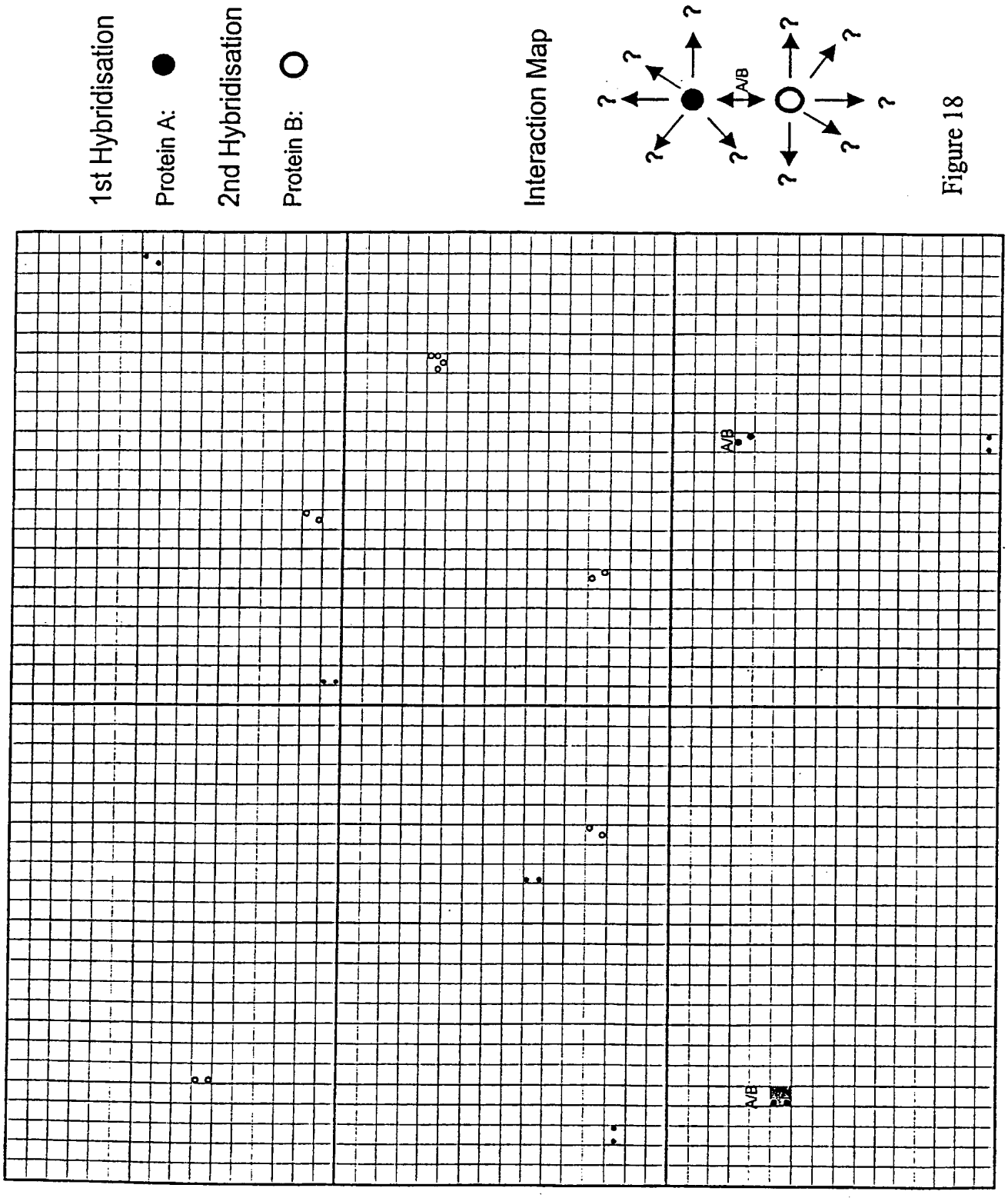

1st Hybridisation
Protein A:
2nd Hybridisation
Protein B:
Interaction Map
A/B
A/B
A/B

Figure 18

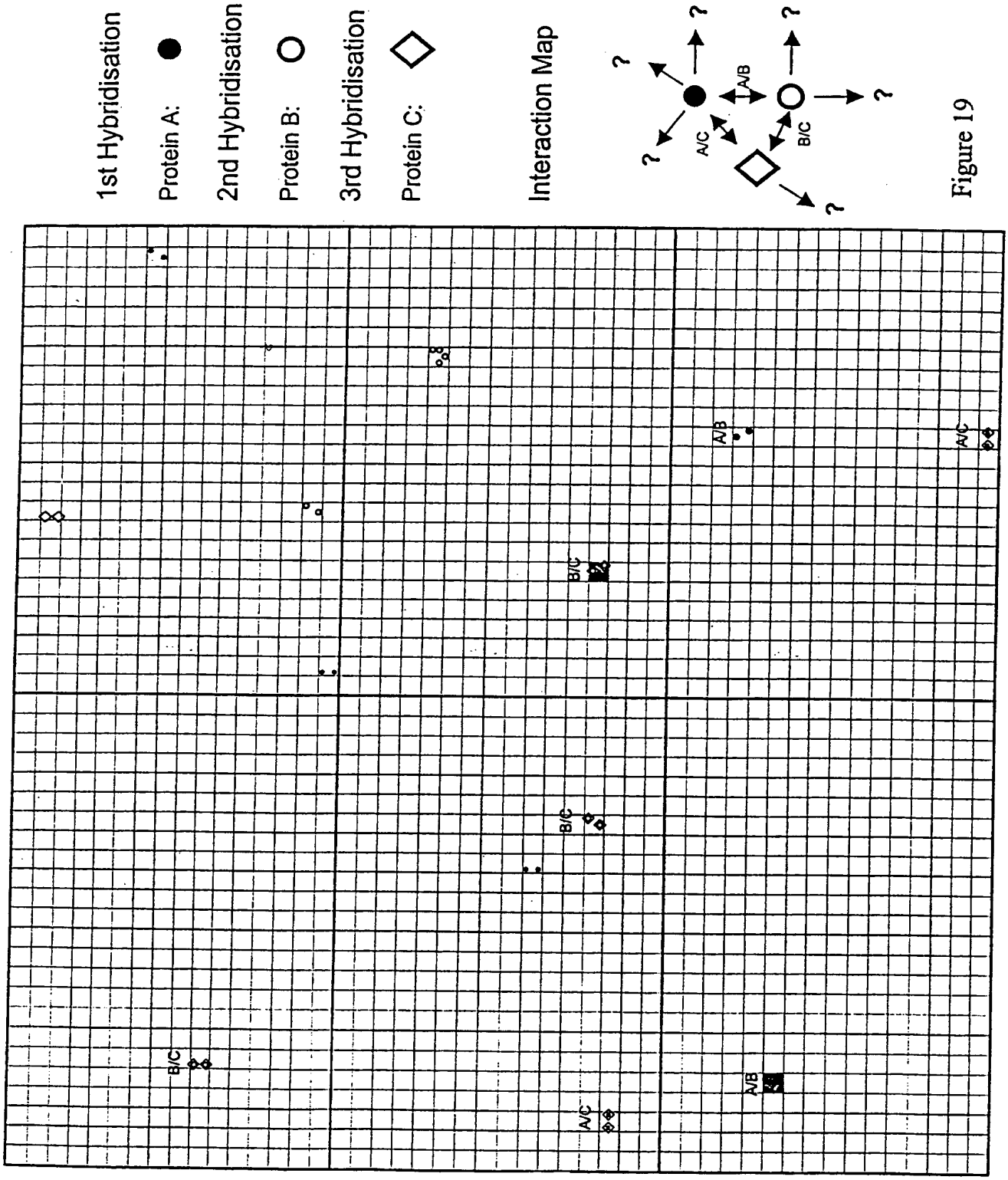


Figure 19

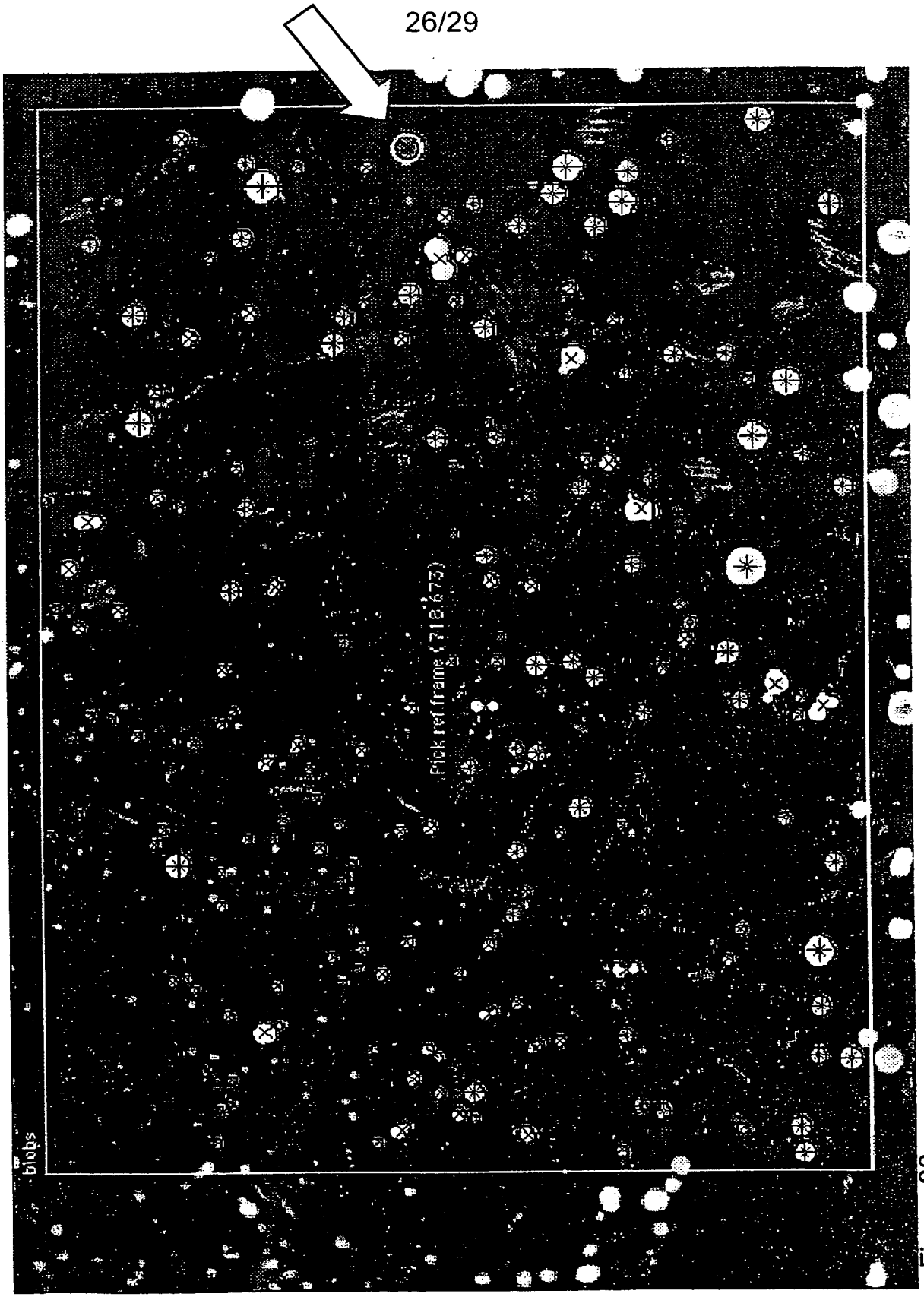


Figure 20

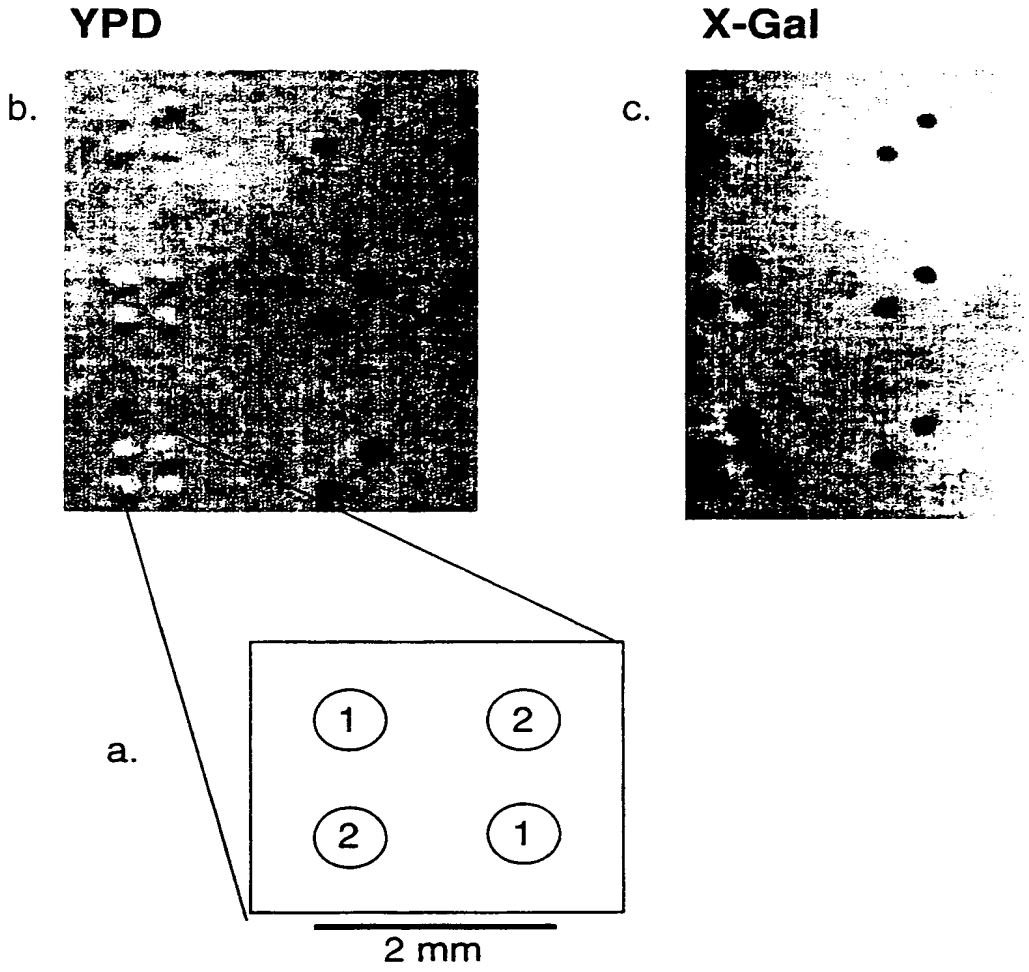


Figure 21

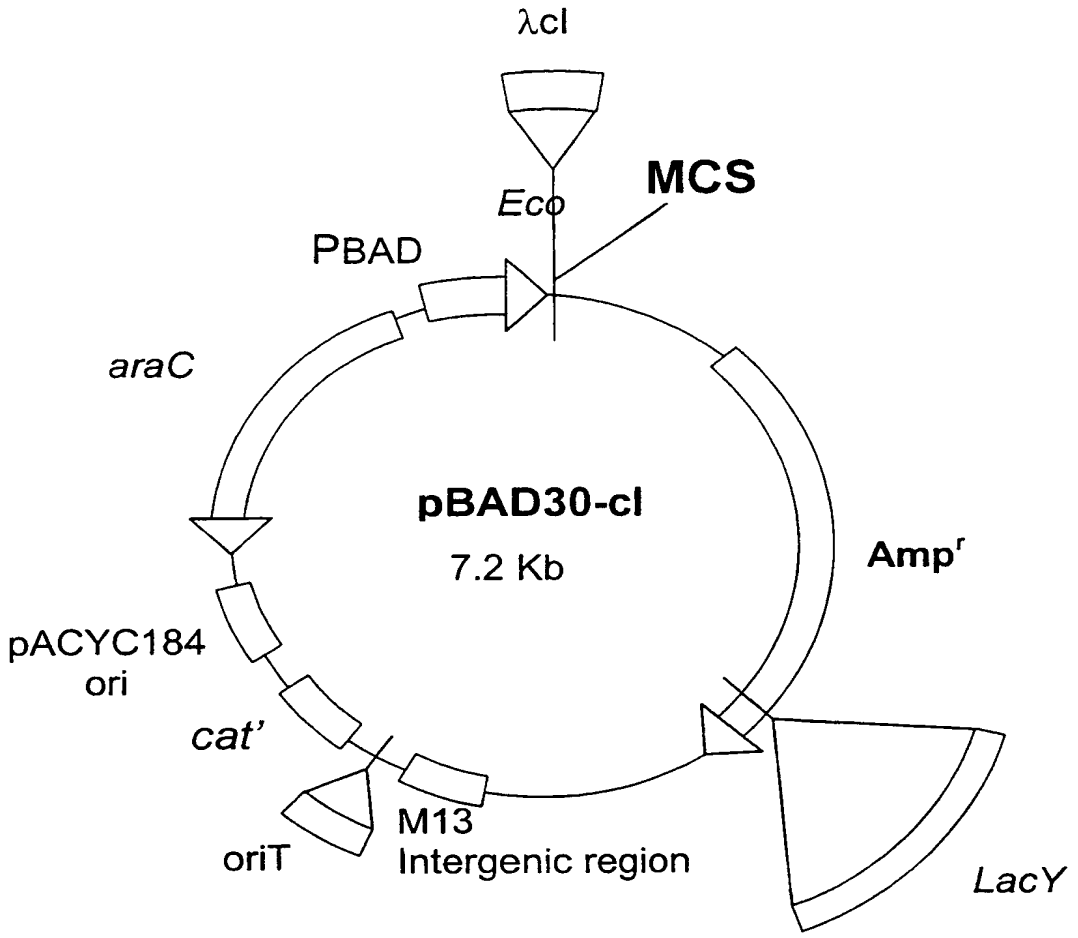

λcI
MCS
Eco
PBAD
araC
pBAD30-cl
7.2 Kb
Amp^r
pACYC184
ori
cat'
M13
Intergenic region
oriT
LacY

Figure 22 (I)

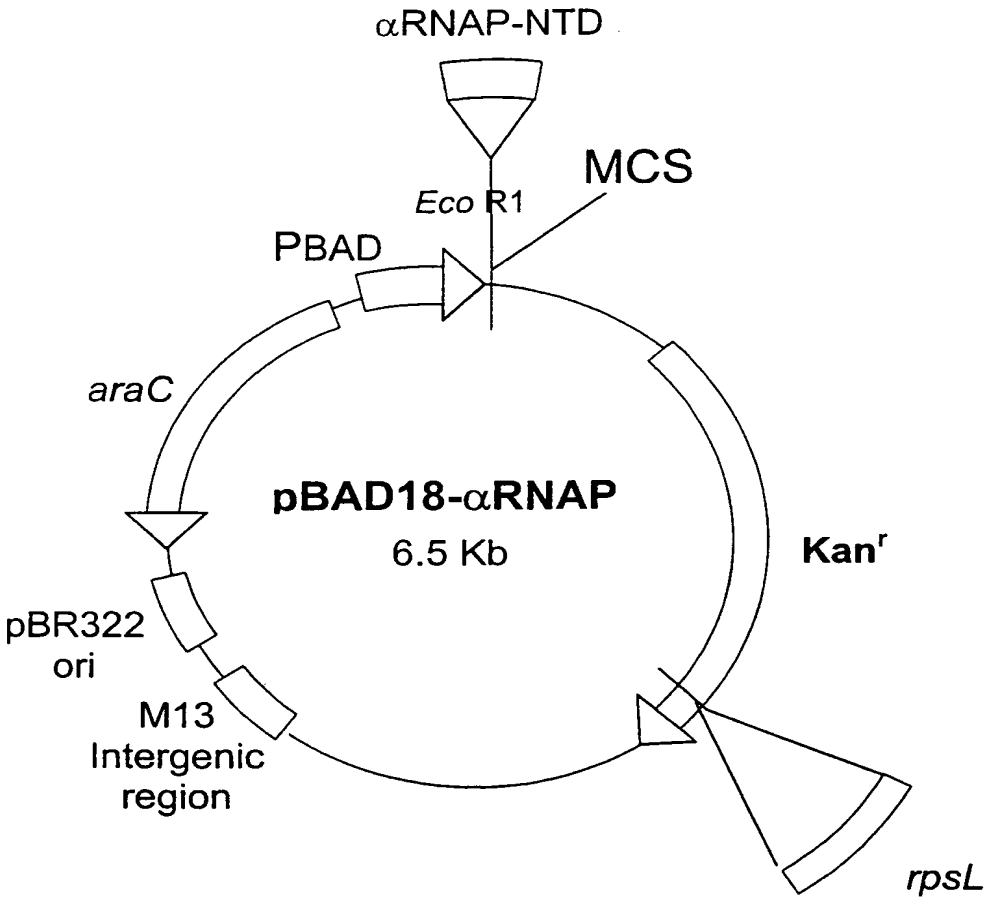


Figure 22 (II)

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/EP 98/07656

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G01N33/50 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G01N C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 32503 A (GEN HOSPITAL CORP ;UNIV JOHNS HOPKINS (US)) 17 October 1996 see page 3, line 10 - page 4, line 9 ---	1-32, 34-36
A	EP 0 790 304 A (AMERSHAM INT PLC) 20 August 1997 see column 11 - column 12 ---	1-32, 34-36
A	DOVE S L ET AL: "ACTIVATION OF PROKARYOTIC TRANSCRIPTION THROUGH ARBITRARY PROTEIN-PROTEIN CONTACTS" NATURE, vol. 386, 10 April 1997, pages 627-630, XP002050708 see the whole document ---	1-32, 34-36

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 April 1999	28/04/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hoekstra, S

3

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/EP 98/07656

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A,P	WO 97 47763 A (CURAGEN CORP) 18 December 1997 see the whole document	1-32, 34-36
A,P	WO 98 07845 A (JOUNG J KEITH ;DOVE SIMON (US); HARVARD COLLEGE (US); HOCHSCHILD A) 26 February 1998 see claims	1-32, 34-36
A	BARTEL, P. ET AL.: "Elimination of false positives that arise in using the two-hybrid system" BIOTECHNIQUES, vol. 14, no. 6, 1993, pages 920-924, XP002004036 see the whole document	1-32, 34-36

INTERNATIONAL SEARCH REPORT

ernational application No.
PCT/EP 98/ 07656

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 31-33, 37-61
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 37-61 have been excluded from the search in view of: Rule 39.1(vi) PCT relating to program for computers; Rule 39.1(v) PCT relating to presentation of information.

2. [X] Claims Nos.: 31-33
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claims Nos.: 31-33

Claims 31-33 are methods for preparing compositions which are technically limited to molecules identified by the screening methods of claims 1-30. No features of a technical nature are comprised in the claims to either characterise the method steps or these molecules. These claims do hence not comply with all the prescribed requirement of Article 6 and Rule 6.3(a) PCt to such an extend that a meaningful search is not possible for these claims.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int ional Application No
PCT/EP 98/07656

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9632503 A	17-10-1996	AU 5541296 A	30-10-1996
		CA 2217545 A	17-10-1996
		EP 0830459 A	25-03-1998
		JP 11502717 T	09-03-1999
EP 0790304 A	20-08-1997	AU 1165697 A	17-07-1997
		EP 0868509 A	07-10-1998
		WO 9723609 A	03-07-1997
WO 9747763 A	18-12-1997	AU 3395597 A	07-01-1998
		CA 2257958 A	18-12-1997
WO 9807845 A	26-02-1998	AU 4159697 A	06-03-1998